



TESIS - SS14 2501

**ESTIMASI MODEL CAMPURAN SPLINE  
*TRUNCATED* DAN KERNEL DALAM REGRESI  
NONPARAMETRIK MULTIVARIABEL**

RISMAL  
NRP. 1314 201 009

DOSEN PEMBIMBING  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. rer. pol. Dedy Dwi Prastyo, M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2016



THESIS - SS14 2501

# **ESTIMATE OF MIXTURE MODEL OF SPLINE TRUNCATED AND KERNEL IN MULTIVARIABLE NONPARAMETRIC REGRESSION**

RISMAL  
NRP. 1314 201 009

SUPERVISORS  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. rer. pol. Dedy Dwi Prastyo, M.Si.

MAGISTER PROGRAM  
DEPARTEMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2016

**ESTIMASI MODEL CAMPURAN SPLINE *TRUNCATED* DAN KERNEL  
DALAM REGRESI NONPARAMETRIK MULTIVARIABEL**

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Sains (M.Si)**

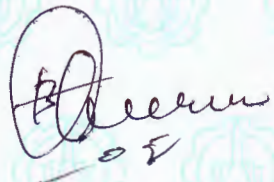
**di  
Institut Teknologi Sepuluh Nopember  
Oleh :**

**RISMAL  
NRP. 1314201009**

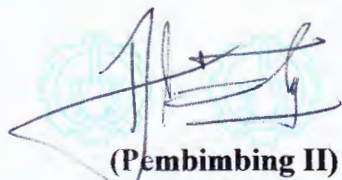
**Tanggal Ujian : 19 Juli 2016  
Periode Wisuda : September 2016**

**Disetujui oleh :**


**1. Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
NIP. 19650603 198903 1 003**

  
(Pembimbing I)

**2. Dr. rer pol. Dedy Dwi Prastyo, S.Si., M.Si.  
NIP. 19831204 200812 1 002**

  
(Pembimbing II)

**3. Dr. Vita Ratnasari, S.Si., M.Si.  
NIP. 19700910 199702 2 001**

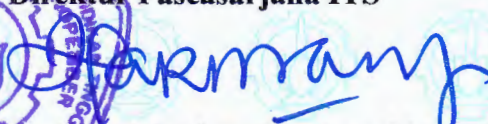
  
(Penguji)

**4. Dr. Ir. Setiawan, MS.  
NIP. 19601030 198701 1 001**

  
(Penguji)



**Direktur Pascasarjana ITS**

  
**Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.**  
NIP. 19601202 198701 1 001

# ESTIMASI MODEL CAMPURAN SPLINE *TRUNCATED* DAN KERNEL DALAM REGRESI NONPARAMETRIK MULTIVARIABEL

Nama : Rismal  
NRP : 1314201009  
Pembimbing : Prof. Dr. I Nyoman Budiantara, M.Si.  
Co-pembimbing : Dr. rer. pol. Dedy Dwi Prastyo, M.Si.

## ABSTRAK

Diberikan data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, y_i)$  dengan variabel prediktor  $(x_{si}, t_{ki})$  dan variabel respon  $y_i$  diasumsikan mengikuti pola data yang tidak diketahui sedemikian sehingga dapat didekati oleh model regresi nonparametrik campuran  $y = \mu(x, t) + \varepsilon = \sum_{s=1}^p f_s(x) + \sum_{k=1}^q g_k(t) + \varepsilon$ . Komponen  $f_s(x)$  didekati dengan fungsi spline aditif dengan prediktor sebanyak  $p$  sementara  $g_k(t)$  didekati dengan fungsi kernel aditif dengan prediktor sebanyak  $q$ . Error  $\varepsilon$  diasumsikan berdistribusi normal dengan mean nol dan varians konstan. Dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) diperoleh estimator  $\hat{f}_s(x) = N(\lambda, \alpha)y$ ,  $\hat{g}_k(t) = M(\alpha)y$ , dan juga  $\hat{\mu}(x, t) = P(\lambda, \alpha)y$ , dimana  $N(\lambda, \alpha), M(\alpha), P(\lambda, \alpha)$  adalah matriks yang bergantung pada knot dan bandwidth. Walaupun estimator yang diperoleh merupakan estimator bias, Namun estimator tersebut merupakan kelas estimator linear dalam observasi  $y$ . Hasil teoritis diterapkan pada data tingkat pengangguran terbuka tahun 2013 di Jawa Barat. Model yang dihasilkan memberikan nilai  $R^2$  sebesar 78.05%.

*Kata Kunci:* Regresi Nonparametrik Campuran, Spline, Kernel, MLE, Tingkat Pengangguran Terbuka.

*(Halaman ini sengaja dikosongkan)*

# ESTIMATE OF MIXTURE MODEL OF SPLINE TRUNCATED AND KERNEL IN MULTIVARIABLE NONPARAMETRIC REGRESSION

Name : Rismal  
Student ID : 1314201009  
Supervisor : Prof. Dr. I Nyoman Budiantara, M.Si.  
Co-supervisor : Dr. rer. pol. Dedy Dwi Prastyo, M.Si.

## ABSTRACT

Given a pair of data  $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, y_i)$  with predictors  $(x_{si}, t_{ki})$  and response variables  $y_i$  are assumed to follow unknown function such that their dependence can be approximated by a mixture nonparametric regression model  $y = \mu(x, t) + \varepsilon = \sum_{s=1}^p f_s(x) + \sum_{k=1}^q g_k(t) + \varepsilon$ . The component  $f_s(x)$  is approximated with additive spline functions with  $p$ -number of predictors whereas  $g_k(t)$  is approximated with additive kernel functions with  $q$ -number of predictors. The error  $\varepsilon$  is assumed normally distributed with mean zero and constant variance. By means of Maximum Likelihood Estimation (MLE) method, it was obtained  $\hat{f}_s(x) = N(\lambda, \alpha)y$ ,  $\hat{g}_k(t) = M(\alpha)y$ , as well as  $\hat{\mu}(x, t) = P(\lambda, \alpha)y$ , where  $N(\lambda, \alpha)$ ,  $M(\alpha)$ ,  $P(\lambda, \alpha)$  are the matrices depending on knots and bandwidths. The estimators obtained was empirically applied on open unemployment rate in West Java in 2013. The resulted model yielded  $R^2$  value 78.05%

**Keywords:** Mixture Nonparametric Regression, Spline, Kernel, MLE, Open Unemployment Rate.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	i
ABSTRAK	iii
ABSTRACT	v
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	7
1.3 Tujuan Penelitian	7
1.4 Manfaat Penelitian	7
1.5 Batasan Permasalahan	8
BAB 2. TINJAUAN PUSTAKA	9
2.1 Regresi Parametrik dan Nonparametrik	9
2.2 Regresi Nonparametrik Kernel	10
2.3 Regresi Nonparametrik Spline	14
2.4 Regresi Campuran Nonparametrik Spline Truncate dan Kernel	15
2.5 Kriteria Kebaikan Model	16
2.6 Tingkat Pengangguran Terbuka	17
BAB 3. METODE PENELITIAN	21
3.1 Sumber Data	21
3.2 Variabel Penelitian	21
3.3 Langkah-langkah Penelitian	23
BAB 4. HASIL DAN PEMBAHASAN	27
4.1 Bentuk Estimasi Model Campuran Regresi Nonparametrik Spline dan Kernel	27
4.2 Sifat Estimator Campuran Regresi dan Kernel	36



4.3 Pemodelan Tingkat Pengangguran Terbuka Propinsi Jawa Barat Menggunakan Campuran Regresi Nonparametrik Spline Truncated dan Kernel	38
4.3.1 Eksplorasi Data	39
4.3.2 Model Umum Regresi Campuran Nonparametrik Spline dan Kernel	41
4.3.3 Pemilihan Banyaknya Titik-titik Knot, dan Lebar <i>Bandwidth</i>	43
4.3.4 Model untuk Tiga Titik Knot	45
4.4 Interpretasi Model dan Analisis Kebijakan	48
BAB 5 KESIMPULAN DAN SARAN	53
5.1 Kesimpulan	53
5.2 Saran	54
DAFTAR PUSTAKA	57
LAMPIRAN	63
BIOGRAFI PENULIS	79

## **DAFTAR GAMBAR**

Gambar	Judul	Halaman
Gambar 4.1	Digram Pencar variabel respon TPT dengan ketujuh prediktor	40

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

Tabel	Judul	Halaman
Tabel 3.1.	Variabel Penelitian	21
Tabel 3.2	Struktur Data Penelitian	23
Tabel 4.1	Statistika Deskriptif TPT dengan 7 Variabel Prediktor	40
Tabel 4.2	Perbandingan Nilai GCV untuk Masing-masing Metode Secara Univariabel	44
Tabel 4.3	Perbandingan Nilai GCV untuk Masing-masing Metode Secara Multivariabel	45
Tabel 4.3	Titik Knot dan GCV untuk Model Campuran	46
Tabel 4.4	Analisis Kebijakan TPT	51

*(Halaman ini sengaja dikosongkan)*

## **DAFTAR LAMPIRAN**

<b>Tabel</b>	<b>Judul</b>	<b>Halaman</b>
Lampiran 1	Data dan Format Data yang digunakan	63
Lampiran 2	Struktur Data pada Matlab	64
Lampiran 3	Fungsi Penentuan Nilai GCV Univariabel (Kernel)	65
Lampiran 4	Fungsi Penentuan Nilai GCV Univariabel (Spline satu knot)	66
Lampiran 5	Fungsi Penentuan Nilai GCV Univariabel (Spline dua knot)	68
Lampiran 6	Fungsi Penentuan Nilai GCV Univariabel (Spline tiga knot)	70
Lampiran 7	Fungsi Penentuan Nilai GCV Multivariabel Empat Spline Tiga Knot dan Tiga Kernel	72
Lampiran 8	Fungsi Penentuan Nilai GCV Multivariabel dua Spline Tiga Knot dan lima Kernel dengan Nilai Knot dan Bandwidth diberikan	74
Lampiran 9	Syntax Untuk Analisis Kebijakan	77

*(Halaman ini sengaja dikosongkan)*

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Penelitian tentang pola hubungan antara variabel respon dan prediktor telah banyak dilakukan. Analisis yang biasa digunakan adalah analisis regresi. Analisis regresi merupakan salah satu metode statistika yang bertujuan untuk memodelkan hubungan antara variabel respon dengan variabel prediktor. Beberapa pendekatan analisis regresi di antaranya adalah regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Regresi parametrik diterapkan apabila bentuk kurva regresi diketahui. Dalam prakteknya, tidak semua data mengikuti pola tertentu sehingga apabila dipaksakan digunakan regresi parametrik, maka akan memberikan kesimpulan yang menyesatkan (Samsodin, 2012). Solusi untuk permasalahan ini adalah regresi nonparametrik.

Regresi nonparametrik adalah pendekatan model yang mengasumsikan pola kurva data tidak diketahui dan termuat di dalam suatu fungsi tertentu, seperti ruang *Sobolev*  $W_2^m[a, b]$ , ruang fungsi-fungsi kontinu  $C(0, \pi)$ , ruang *Hilbert*, ruang *entropi* dan yang lainnya (Budiantara, 2000). Oleh karena itu dengan menggunakan pendekatan ini model yang didapatkan akan lebih baik karena data diharapkan mencari sendiri bentuk estimasi modelnya tanpa dipengaruhi subyektifitas oleh peneliti dan juga pendekatan ini sangat fleksibel (Eubank, 1988). Regresi semiparametrik adalah regresi dimana pola data variabel-variabel prediktornya terhadap variabel respon berpola gabungan antara parametrik dan nonparametrik atau dengan kata lain regresi semiparametrik digunakan jika salah satu kurva regresinya tidak diketahui sedangkan yang lainnya diketahui (Eubank, 1999 dan Budiantara, 2009). Penelitian tentang regresi semiparametrik juga telah banyak dilakukan oleh peneliti.

Dalam beberapa tahun terakhir, seiring dengan berkembangnya komputasi dan beberapa keterbatasan pada model regresi parametrik, maka model regresi nonparametrik yang tidak memerlukan banyak asumsi menjadi lebih banyak diaplikasikan untuk menyelesaikan suatu permasalahan di berbagai bidang terapan



(Mahler, 1995). Model regresi nonparametrik dapat digunakan dengan tujuan kebijakan, prediksi, dan peramalan.

Ada beberapa teknik estimasi dalam regresi nonparametrik, antara lain: estimator histogram, *spline*, *kernel*, k-NN, *orthogonal*, *fourier*, dan *wavelet* (Eubank, 1999). Beberapa penelitian dasar dan pengembangan yang berkaitan dengan estimator-estimator tersebut dapat dilihat pada Manzan (2005), Okumura dan Naito (2006), You dan Chen (2007), Yao (2007), Kayri dan Zirbhoglu (2009), Chobey (2012) untuk estimator kernel. Craven dan Wahba (1979), Wahba (1990), Green dan Silverman (1994), Eubank (1988), Liang (2006), Lin dan Zhang (2006), Budiantara *et al.* (2010), Ma dan Yang (2011), Montoril (2014) untuk estimator spline. Welsh dan Yee (2006), Su dan Ullah (2008), He dan Huang (2009), Martins-Filho dan Yao (2009), serta Qingguo (2010) untuk estimator polynomial local. Antoniadis, *et al.* (2001), Amato dan De-Canditiis (2001), Li (2002), Angelini, *et al.* (2003), Qu (2002), Rakotomamonjy, *et al.* (2005), serta Taylor (2009) untuk estimator wavelet. Dan De Jong (1977), Bilodeau (1992), Amato dan De-Canditiis (2001), Oirrak (2001), Amato, *et al.* (2002), Tripena dan Budiantara (2007), serta Galtchouk dan Pergamenshchikov (2009) untuk estimator deret fourier.

Salah satu estimator yang sering digunakan dalam regresi nonparametrik yaitu fungsi kernel. Ada beberapa jenis fungsi kernel, antara lain kernel *uniform*, *triangle*, *epanechnikov*, *gaussian*, *kuartik* dan *cosinus* (Hardle, 1994). Untuk mendapatkan estimator kernel tersebut, metodenya dinamakan pendekatan *direct*. Pada pendekatan *direct*, pemilihan parameter pemulus (*bandwidth*) jauh lebih penting dibandingkan dengan memilih fungsi kernel itu sendiri. Sehingga yang menjadi permasalahan dalam regresi kernel adalah pemilihan *bandwidth*, bukan pada pemilihan fungsi kernel. Fungsi kernel yang umum digunakan adalah kernel gaussian dan kernel Epanechnikov (Hastie dan Tibshirani, 1990).

Pendekatan dengan fungsi kernel memiliki bentuk yang lebih fleksibel, bentuk maupun perhitungan matematisnya mudah dan lebih sederhana, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara dan Mulianah, 2007, Fan dan Liu, 2013). Keuntungan lain dalam menggunakan fungsi kernel adalah kemudahan dalam penghitungan sifat dasar estimator seperti bias dan

varians, serta didapatkan jawaban andal dan sederhana untuk pengembangan aplikasi yang lebih luas (Sonhaji, 2006). Jawaban andal maksudnya bahwa jawaban yang diperoleh lebih mendekati sebaran data yang sebenarnya sehingga meminimumkan residual. Sedangkan sederhana maksudnya bahwa proses penghitungan dalam regresi nonparametrik kernel lebih mudah dimengerti.

Spline merupakan salah satu model yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik (Eubank, 1988). Disamping itu spline mampu menangani karakter data/fungsi yang bersifat mulus (*smooth*). Spline juga memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Cox dan O'Sullivan, 1996). Salah satu jenis estimator spline yang paling sering digunakan regresi spline *truncated*. Pendekatan regresi spline *truncated* mempunyai beberapa kelebihan diantaranya adalah lebih mudah secara matematis dan interpretasi hampir sama seperti pada regresi parametrik. Hal ini dikarenakan spline merupakan salah satu jenis potongan polinomial yang memiliki sifat tersegmen. Dengan memiliki sifat tersegmen maka mengakibatkan tingkat fleksibilitas yang lebih tinggi dibandingkan dengan jenis potongan polinomial biasa. Besar kemungkinan dengan sifat polinomial tersegmen ini menjadikan model spline mampu menyesuaikan diri secara lebih efektif terhadap karakteristik lokal pada suatu fungsi atau data. Dalam proses estimasi, apabila estimator spline yang diperoleh berdasarkan optimasi *Penalized Least Square* (PLS), maka persoalan utama dalam estimator ini adalah pemilihan parameter penghalus yang optimal, sedangkan apabila estimator spline yang diperoleh dengan optimasi *Least Square* (LS), maka persoalan utama dalam estimator ini adalah pemilihan titik-titik knot yang optimal.

Terdapat tiga kriteria yang harus diperhatikan dalam membentuk model regresi spline. Kriteria yang pertama adalah orde untuk model, yang dapat ditentukan berdasarkan pola umum yang terjadi pada data. Kriteria yang kedua menentukan jumlah knot, dimana penentuan banyaknya knot berdasarkan jumlah perubahan pola di daerah tertentu pada kurva. Dimana setiap terjadi perubahan pola, maka disitu terdapat titik knot, sehingga banyaknya titik knot dapat ditentukan. Knot dapat diartikan sebagai titik perpaduan bersama dimana terdapat

perubahan perilaku pola pada interval yang berlainan. Kriteria yang ketiga adalah menentukan lokasi knot, hampir sama dengan kriteria yang kedua, untuk menentukan lokasi knot dilakukan dengan berdasarkan perubahan pola di daerah tertentu pada kurva.

Estimator lain yang biasa digunakan adalah polinomial lokal. Salah satu kelebihan estimator polinomial lokal adalah dapat mengurangi asimtotik bias dan menghasilkan estimasi yang baik. Estimator polinomial lokal dapat diperoleh dengan optimasi WLS (*Weighted Least Squared*), sedangkan untuk mengestimasi parameter penghalus (*bandwidth*) dilakukan dengan menggunakan metode *Generalized Cross Validation* (GCV). Beberapa penelitian tentang polinomial lokal diberikan oleh Welsh dan Yee (2006) yang mengkaji tentang penggunaan polinomial lokal pada regresi nonparametrik dengan variabel respon lebih dari satu. Pada kajiannya, Welsh dan Yee (2006) telah menurunkan sifat bias dan varians asimtotik dari estimator polinomial lokal tersebut. Estimator yang sekarang banyak dikembangkan adalah deret fourier dan wavelet. De Jong (1977) merupakan kelompok peneliti awal yang mengkaji tentang transformasi deret fourier untuk penghalusan fungsi densitas, khususnya pada penghalusan estimator spektral. Pada tahun 1992. Kajian tentang wavelet pada regresi nonparametrik antara lain dapat dilihat pada Antoniadis *et al.* (2001) yang membahas tentang metode wavelet untuk memodelkan observasi dan suatu signal yang terkontaminasi oleh gangguan (*noise*) yang berdistribusi gauss dan bersifat aditif.

Penelitian regresi nonparametrik saat ini telah banyak melibatkan prediktor. Penelitian tersebut pada umumnya menggunakan jenis metode estimasi yang sama untuk sebagian bahkan semua prediktornya. Padahal kemungkinan besar variabel-variabel tersebut memiliki pola data atau kurva regresi yang berbeda. Banyaknya titik knot dan lebar bandwidth juga sering diasumsikan sama, padahal kemungkinan hal tersebut terjadi sangatlah kecil. Pada penelitian ini akan digunakan model campuran dua estimator nonparametrik yang berbeda. Penelitian terdahulu yang melibatkan dua jenis estimator nonparametrik berbeda adalah penelitian dilakukan oleh Sudiarsa (2015) dan Budiantara *et al.* (2015). Penelitian yang dilakukan Sudiarsa (2015) melibatkan estimator deret fourier dan spline

*truncated*. Sedangkan penelitian yang dilakukan Budiantara *et al.* (2015) telah melibatkan kernel dan spline *truncated*.

Berdasarkan uraian tersebut dan mempertimbangkan kelebihan setiap estimator, dalam penelitian ini akan dilakukan pengembangan estimator regresi nonparametrik dengan menggunakan dua jenis estimator nonparametrik untuk seluruh prediktor. Fokus penelitian ditekankan pada model regresi nonparametrik yang aditif dengan gabungan antara kernel dan spline *Truncated* yang diperoleh melalui optimasi *Least Squares (LS)*. Penelitian ini juga mengembangkan estimator spline untuk banyaknya titik knot yang berbeda untuk setiap variabel yang menggunakan estimator spline dan juga lebar bandwidth yang kemungkinan berbeda untuk variabel yang menggunakan estimator kernel. Hasil penelitian teoritis akan diaplikasikan pada studi empiris pada suatu permasalahan nyata, yaitu pemodelan Tingkat Pengangguran Terbuka Provinsi Jawa Barat tahun 2013.

Masalah pengangguran merupakan salah satu permasalahan utama yang harus segera ditangani oleh Pemerintah. Tingkat pengangguran yang tinggi akan menimbulkan berbagai permasalahan sosial. Masalah pengangguran juga menyebabkan tingkat pendapatan nasional dan tingkat kemakmuran masyarakat tidak mencapai potensi maksimal. Tingginya angka pengangguran di Indonesia disebabkan karena jumlah penduduk yang semakin meningkat tidak diimbangi dengan pertumbuhan lapangan usaha yang ada. Indikator utama yang digunakan untuk mengukur angka pengangguran dalam angkatan kerja yaitu Tingkat Pengangguran Terbuka (BPS, 2009). Tingkat Pengangguran Terbuka (TPT) menunjukkan persentase jumlah pencari kerja terhadap jumlah angkatan kerja. Jawa Barat merupakan provinsi dengan Tingkat Pengangguran Terbuka (TPT) yang sangat tinggi dalam beberapa tahun terakhir. Pada Bulan Februari 2013 Badan Pusat Statistik menempatkan Provinsi Jawa Barat sebagai provinsi dengan Tingkat Pengangguran Terbuka (TPT) tertinggi ketiga, setelah Banten dan DKI Jakarta, yakni sebesar 8,90 persen. Peringkat ini tidak mengalami perbaikan sejak setahun lalu (BPS, 2013). Pada Februari 2013, jumlah angkatan kerja di seluruh provinsi Jawa Barat sebanyak 20.388.637 orang dimana penduduk yang bekerja tercatat sebanyak 18.573.371 orang dan yang menganggur sebanyak 1.815.266 orang (BPS, 2013).

Penelitian sebelumnya mengenai Tingkat Pengangguran Terbuka pernah dilakukan oleh Santoso (2009) dengan melakukan klasifikasi Kabupaten/Kota di Jawa Timur berdasarkan Tingkat Pengangguran Terbuka dengan Pendekatan MARS (Multivariate Adaptive Regression Spline). Sementara Asti (2011) meneliti tentang Regresi Multivariate pada faktor yang berpengaruh terhadap pengangguran terbuka di Jawa Timur. Ariane (2012) melakukan pendekatan Regresi Ridge untuk memodelkan faktor-faktor yang mempengaruhi Tingkat Pengangguran Terbuka Di Provinsi Jawa Timur dan Jawa Tengah. Tingkat Pengangguran Terbuka di Jawa Barat pernah diteliti oleh Prihatiningsih (2012) dengan pendekatan Regresi Terboboti Geografis (RTG). Prihatiningsih menunjukkan bahwa faktor yang mempengaruhi TPT di Jawa Barat yaitu kepadatan penduduk, persentase penduduk miskin, upah minimum Kabupaten/Kota, dan persentase unit usaha industri besar per jumlah penduduk usia kerja. Faidah (2012) meneliti Tingkat Pengangguran Terbuka (TPT) perempuan dengan model tobit spasial dan menyatakan bahwa faktor yang mempengaruhi TPT perempuan di Pulau Jawa adalah antara lain: persentase penduduk yang tinggal di daerah perkotaan, seks rasio, persentase penduduk yang berpendidikan diatas SLTP, persentase penduduk yang mampu membaca dan menulis, serta tingkat pertumbuhan ekonomi. Kemudian Anwar (2014) meneliti tentang Tingkat Pengangguran Terbuka di Jawa Barat dengan menggunakan metode analisis regresi *spline*.

Berdasarkan penjelasan diatas, masih belum terdapat penelitian yang meneliti tentang Tingkat Pengangguran Terbuka dengan menggunakan dua jenis metode estimasi nonparametrik. Padahal kemungkinan pola data variabel-variabel prediktor terhadap variabel respon memiliki jenis pola data yang tidak sama. Oleh karena itu, dalam penelitian ini akan diselidiki faktor-faktor yang mempengaruhi Tingkat Pengangguran Terbuka di Jawa Barat dengan menggunakan dua pendekatan estimator nonparametrik, yaitu *spline truncated* dan kernel. Dari hasil yang diperoleh diharapkan dapat digunakan orang-orang yang bekerja di bidang pemerintahan dalam mengambil keputusan.

## 1.2 Perumusan Masalah

Beberapa permasalahan yang ingin dikaji dalam penelitian ini adalah :

1. Bagaimana estimasi model campuran regresi nonparametrik multivariabel kernel dan spline *truncated* dalam regresi nonparametrik multivariabel?
2. Bagaimana sifat-sifat estimator campuran regresi spline *truncated* dan kernel?
3. Bagaimana estimasi model campuran regresi nonparametrik kernel dan spline *truncated* pada data Tingkat Pengangguran Terbuka Provinsi Jawa Barat tahun 2013?

## 1.3 Tujuan Penelitian

Berdasarkan perumusan masalah tersebut di atas, tujuan penelitian ini adalah:

1. Mendapatkan estimator campuran kernel dan spline *truncated* dalam regresi nonparametrik multivariabel.
2. Mendapatkan sifat-sifat estimator campuran regresi spline *truncated* dan kernel
3. Memodelkan Tingkat Pengangguran Terbuka Provinsi Jawa Barat tahun 2013 menggunakan estimator campuran kernel dan spline *truncated* dalam regresi nonparametrik multivariabel.

## 1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dari penelitian ini adalah sebagai berikut :

1. Secara teoritis diharapkan dapat memberikan wawasan keilmuan yang lebih khusus kepada penulis tentang model campuran regresi nonparametrik multivariabel kernel dan spline *truncated*.
2. Mengetahui penerapan model campuran regresi nonparametrik multivariabel kernel dan spline *truncated* pada kasus Tingkat Pengangguran Terbuka Provinsi Jawa Barat tahun 2013.

## 1.5 Batasan Masalah

Mengacu perumusan masalah di atas, maka ruang lingkup dalam penelitian ini dibatasi pada beberapa hal, antara lain sebagai berikut:

1. Spline yang digunakan pada penelitian ini adalah spline *truncated* dengan menggunakan GCV untuk pemilihan titik knot optimumnya.
2. Fungsi kernel yang digunakan pada penelitian ini adalah kernel *gaussian*.

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Regresi Parametrik dan Nonparametrik

Pendekatan regresi parametrik digunakan jika penyebaran pola data membentuk pola tertentu, apakah linear, kuadratik atau kubik. Untuk dapat menggunakan model regresi parametrik diperlukan pengalaman masa lalu atau terdapat sumber-sumber lain yang tersedia dan dapat memberikan informasi yang detail tentang data. Dalam regresi parametrik terdapat asumsi tegas, yaitu bentuk kurva regresi diketahui (Eubank, 1999). Metode digunakan untuk estimasi parameter adalah *Least Square* dan *Maximum Likelihood* (Wahba, 1990). Model Pendekatan yang dihasilkan oleh metode-metode tersebut ekuivalen dengan estimasi parameter dalam model. Secara umum regresi parametrik linear dapat ditulis dalam bentuk (Montgomery dan Peak, 1991)

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_r x_{ri} + \varepsilon_i, i = 1, 2, \dots, n, \quad (2.1)$$

dimana  $y$  variabel respon,  $x_1, x_2, \dots, x_r$  sebagai variabel prediktor,  $\varepsilon_i$  merupakan *error* random independen berdistribusi normal dengan mean nol dan variansi  $\sigma^2$  parameter  $\beta_0, \beta_1, \beta_2, \dots, \beta_r$  tidak diketahui. Jika model yang diasumsikan ini benar maka estimasi regresi parametrik sangat efisien, tetapi jika tidak akan menyebabkan interpretasi data yang menyesatkan (Simonoff, 1996).

Secara umum bentuk regresi parametrik multivariabel dengan  $r$  variabel prediktor pada persamaan (2.1) dapat ditulis dalam bentuk matriks:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.2)$$

dengan,  $\mathbf{y}$  vektor respon yang berukuran  $n \times 1$ ,  $\mathbf{X}$  matrik berukuran  $n \times (r + 1)$ ,  $\boldsymbol{\beta}$  vektor parameter yang akan diestimasi berukuran  $(r + 1) \times 1$ , vektor *error*  $\boldsymbol{\varepsilon}$  random berukuran  $n \times 1$ ,  $\varepsilon_i$  berdistribusi  $N(0, \sigma^2)$ . Bentuk vektor dan matriks pada persamaan (2.2) adalah (Ryan, 1997):



$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1r} \\ 1 & x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nr} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix}, \text{ dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_r \end{pmatrix}$$

Parameter  $\boldsymbol{\beta}$  diestimasi dengan menggunakan metode *Ordinary Least Square* (OLS) yang meminimumkan  $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$  dimana  $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ . Dengan menurunkan parsial  $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$  terhadap  $\boldsymbol{\beta}$  dan menyamakan dengan nol, maka diperoleh estimator  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$

Misalkan diberikan data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{qi}, y_i)$ , hubungan antara  $(x_{1i}, x_{2i}, \dots, x_{qi})$  dengan  $y_i$  dimodelkan dengan regresi nonparametrik multivariabel aditif:

$$y_i = f_1(x_{1i}) + f_2(x_{2i}) + \cdots + f_q(x_{qi}) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.3)$$

Error  $\varepsilon_i$  diasumsikan berdistribusi normal independen dengan mean nol dan varian  $\sigma^2$ . Bentuk kurva regresi  $f_k$  dengan  $k = 1, 2, \dots, q$  diasumsikan tidak diketahui. Bentuk model (2.3) dapat ditulis dalam vektor sebagai berikut:

$$\mathbf{y} = \sum_{k=1}^q \mathbf{f}_k(x_k) + \boldsymbol{\varepsilon},$$

dengan  $\mathbf{y} = (y_1 \cdots y_n)'$ ,  $\mathbf{f}_k(x_k) = (f_k(x_{k1}) \cdots f_k(x_{kn}))'$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1 \cdots \varepsilon_n)'$ .

Kurva regresi nonparametrik  $f_k(x_k)$ ,  $k = 1, 2, \dots, q$  diestimasi dengan metode *least square* (Wang, 2011).

## 2.2 Regresi Nonparametrik Kernel

Suatu fungsi  $K : R \rightarrow R$  disebut fungsi kernel jika  $K$  fungsi kontinu, simetris, terbatas dan

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad (2.4)$$

Dari definisi tersebut, jika  $K$  adalah fungsi nonnegatif maka  $K$  juga diartikan sebagai suatu fungsi padat peluang (fungsi densitas). Secara umum, kernel  $K$  dengan *bandwidth*  $\alpha$  didefinisikan oleh:

$$K_{\alpha}(x) = \frac{1}{\alpha} K\left(\frac{x}{\alpha}\right) \quad -\infty < x < \infty, \quad \alpha > 0 \quad (2.5)$$

serta memenuhi

- $K(x) \geq 0$  , untuk semua  $x$
- $\int_{-\infty}^{\infty} K(x) dx = 1$
- $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$
- $\int_{-\infty}^{\infty} x K(x) dx = 0$

(Wand dan Jones, 1995)

Beberapa jenis fungsi kernel (Hardle, 1990) antara lain:

- (1) *kernel uniform*:  $K(x) = \frac{1}{2} I(|x| \leq 1)$
- (2) *kernel triangle*:  $K(x) = (1 - |x|) I(|x| \leq 1)$
- (3) *kernel epanechnikov*:  $K(x) = \frac{3}{4} (1 - x^2) I(|x| \leq 1)$
- (4) *kernel kuartik*:  $K(x) = \frac{15}{16} (1 - x^2)^2 I(|x| \leq 1)$
- (5) *kernel triweight*:  $K(x) = \frac{35}{32} (1 - x^2)^3 I(|x| \leq 1)$
- (6) *kernel cosinus*:  $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2} x\right) I(|x| \leq 1)$
- (7) *kernel gaussian*:  $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty,$

dengan  $I$  adalah fungsi indikator. Penduga densitas kernel untuk fungsi densitas  $g(t)$  adalah

$$\hat{g}_{\alpha}(t) = \frac{1}{n} \sum_{i=1}^n K_{\alpha}(t - t_i) = \frac{1}{n\alpha} \sum_{i=1}^n K\left(\frac{t - t_i}{\alpha}\right) \quad (2.6)$$

Pada persamaan (2.6) terlihat bahwa  $\hat{g}_\alpha(t)$  tergantung pada fungsi kernel  $K$  dan parameter  $\alpha$ . Bentuk bobot kernel ditentukan oleh fungsi kernel  $K$ , sedangkan ukuran bobotnya ditentukan oleh parameter pemulus  $\alpha$  yang disebut *bandwidth*.

Regresi kernel adalah teknik statistika nonparametrik untuk menduga fungsi regresi  $f(t_i)$  pada model regresi nonparametrik:

$$y_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.7)$$

Secara teoritis fungsi regresi kernel (Carmona, 2003) didefinisikan sebagai:

$$f(t) = E(Y|T=t) = \int_{-\infty}^{\infty} y g(y|t) dy = \frac{\int_{-\infty}^{\infty} y g(t, y) dy}{g(t)} \quad (2.8)$$

Fungsi polinomial lokal mengadopsi dari perluasan deret Taylor di sekitar  $v$ . Jika  $h(v_i)$  pada persamaan (2.7) dapat dihampiri dengan fungsi polinomial lokal:

$$\begin{aligned} f(t_i) &= \mathcal{G}_0 + \mathcal{G}_1(t_i - t) + \mathcal{G}_2(t_i - t)^2 + \dots + \mathcal{G}_p(t_i - t)^p \\ &= \sum_{k=0}^p \mathcal{G}_k(t_i - t)^k, \end{aligned}$$

dimana:

$$\mathcal{G}_k = \frac{h^k(t)}{k!}, \text{ untuk } k = 1, 2, \dots, n.$$

Dalam bentuk matriks dapat ditulis:

$$\mathbf{y} = \mathbf{T}\mathcal{G} + \boldsymbol{\varepsilon},$$

dimana:

$$\mathbf{T} = \begin{bmatrix} 1 & (t_1 - t) & (t_1 - t)^2 & \dots & (t_1 - t)^p \\ 1 & (t_2 - t) & (t_2 - t)^2 & \dots & (t_2 - t)^p \\ & \vdots & \vdots & \ddots & \vdots \\ 1 & (t_n - t) & (t_n - t)^2 & \dots & (t_n - t)^p \end{bmatrix}, \mathcal{G} = \begin{bmatrix} \mathcal{G}_0 \\ \mathcal{G}_1 \\ \mathcal{G}_2 \\ \vdots \\ \mathcal{G}_p \end{bmatrix}$$

Parameter  $\mathcal{G}$  bergantung pada titik  $v$  yang disebut sebagai titik lokal. Parameter  $\mathcal{G}$  dapat diestimasi menggunakan *weighted least square* (WLS) dengan meminimumkan:

$$L = \sum_{i=1}^n \left( y_i - \sum_{k=0}^p \mathcal{G}_k (t_i - t)^k \right)^2 K \left( \frac{t_i - t}{\alpha} \right), \quad (2.9)$$

dimana  $\alpha$  merupakan bandwidth yang mengontrol ukuran persekitaran titik lokal  $t_0$ . Permasalahan *weighted least square* pada (2.9) dalam bentuk matriks dapat ditulis menjadi:

$$\mathbf{L} = (\mathbf{y} - \mathbf{T}\mathcal{G})'\mathbf{W}(\mathbf{y} - \mathbf{T}\mathcal{G})$$

dimana  $\mathbf{W}$  merupakan matriks diagonal yang berukuran  $n \times n$ .

$$W = \text{diag} \left( K \left( \frac{t_i - t}{\alpha} \right) \right); i = 1, 2, \dots, n$$

Hasil estimasi  $\mathcal{G}$  adalah:

$$\hat{\mathcal{G}} = (\mathbf{T}'\mathbf{W}\mathbf{T})^{-1}\mathbf{T}'\mathbf{W}\mathbf{y}$$

sehingga estimasi model regresi polinomial lokal adalah

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{T}(\mathbf{T}'\mathbf{W}\mathbf{T})^{-1}\mathbf{T}'\mathbf{W}\mathbf{y} \\ &= \mathbf{H}\mathbf{y} \end{aligned}$$

Terkait dengan model regresi polinomial lokal, model regresi kernel Nadaraya-Watson merupakan model regresi polinomial lokal yang memuat konstan lokal. Jadi jika  $f(t_i)$  hanya memuat konstan lokal, maka dengan meminimumkan:

$$L = \sum_{i=1}^n (y_i - \mathcal{G}_0)^2 K \left( \frac{t_i - t}{\alpha} \right)$$

akan dihasilkan:

$$\mathcal{G}_0 = \sum_{i=1}^n \frac{K \left( \frac{t_i - t}{\alpha} \right)}{\sum_{i=1}^n K \left( \frac{t_i - t}{\alpha} \right)} y_i$$

sehingga:

$$\begin{aligned} \hat{f}(t) &= n^{-1} \sum_{i=1}^n \left( \frac{K_{\alpha}(t - t_i)}{n^{-1} \sum_{i=1}^n K_{\alpha}(t - t_i)} \right) y_i \\ &= \sum_{i=1}^n W_{\alpha i}(t) y_i, \end{aligned} \quad (2.10)$$

dengan

$$W_{ai}(t) = \frac{n^{-1}K_{\alpha}(t-t_i)}{n^{-1}\sum_{i=1}^n K_{\alpha}(t-t_i)} = \frac{K\left(\frac{t-t_i}{\alpha}\right)}{\sum_{i=1}^n K\left(\frac{t-t_i}{\alpha}\right)} \quad (2.11)$$

Penduga (2.10) diusulkan oleh Nadaraya dan Watson, sehingga penduga ini sering disebut penduga Nadaraya-Watson (Hardle, 1994).

### 2.3 Regresi Nonparametrik Spline

Potongan polinomial memainkan peranan penting dalam teori aproksimasi dan stalistika. Potongan polinomial mempunyai sifat fleksibel dan efektif untuk menangani sifat lokal suatu fungsi atau data (Eubank, 1999). Salah satu jenis potongan polinomial yang penting adalah polinomial spline. Spline sebagai pendekatan pola data dikenakan oleh Whittaker pada tahun 1923, sedangkan spline yang didasarkan pada suatu persoalan optimasi dikembangkan oleh Reinsch pada tahun 1967 (Wahba, 1990).

Polinomial spline dengan ordo  $m$  dengan titik-titik knot  $\lambda_1, \lambda_2, \dots, \lambda_r$  ( $a < \lambda_1 < \dots < \lambda_r < b$ ) adalah fungsi  $f$  yang disajikan dalam bentuk:

$$f(x) = \sum_{i=0}^m \theta_i x^i + \sum_{j=1}^r \phi_j (x - \rho_j)_+^m, \quad (2.12)$$

dengan  $\theta_i, i = 0, 1, \dots, m$  dan  $\phi_j, j = 1, 2, \dots, r$  adalah konstanta-konstanta bernilai real, serta:

$$(x - \lambda_j)_+^m = \begin{cases} (x - \lambda_j)^m & , \text{jika } x \geq \lambda_j \\ 0 & , \text{jika } x < \lambda_j \end{cases}$$

Penelitian estimator spline pada model regresi nonparamctrik telah banyak dilakukan seperti Huang (2003), Crainieeanu, *et al.* (2004), Kirn dan Gu (2004), Lee (2004), Budiantara (2010), serta Howell (2007). Model regresi nonparamctrik yang menyatakan hubungan antara variabel prediktor dengan variabel respon dapat ditulis sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.13)$$

dengan

$x_i$  : variabel prediktor untuk pengamatan ke- $i$  untuk  $i = 1, 2, \dots, n$   
 $y_i$  : variabel respon ke- $i$  untuk  $i = 1, 2, \dots, n$   
 $\varepsilon_i$  : variabel error (random) pada pengamatan ke- $i$ , dengan  $i = 1, 2, \dots, n$

Persamaan (2.12) dapat diubah dalam bentuk perkalian matrik

$$\mathbf{f}(x) = \mathbf{X}(\lambda)\boldsymbol{\beta}, \quad (2.14)$$

dengan  $\mathbf{X}(\lambda)$  adalah matrik yang bergantung pada  $\lambda$ , sedangkan  $\boldsymbol{\beta}$  merupakan vektor yang bergantung pada  $\boldsymbol{\theta}$ , dan  $\phi$ . Untuk mendapatkan estimator dari  $f$ , juga digunakan metode Maximum Likelihood Estimation (MLE). Estimasi yang pertama dilakukan adalah mengestimasi  $\boldsymbol{\beta}$ , dengan fungsi likelihood:

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2 | \lambda) &= \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2}} \text{Exp} \left( -\frac{\varepsilon_i^2}{2\sigma^2} \right) \right) \\ &= (2\pi\sigma^2)^{-n/2} \text{Exp} \left( -\frac{1}{2\sigma^2} \|\boldsymbol{\varepsilon}\|^2 \right) \\ &= (2\pi\sigma^2)^{-n/2} \text{Exp} \left( -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta}\|^2 \right) \end{aligned}$$

sehingga diperoleh:

$$\hat{\boldsymbol{\beta}}(\lambda) = \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' \mathbf{y}$$

Menyebabkan estimasi untuk  $\mathbf{f}(x)$  adalah:

$$\hat{\mathbf{f}}(x) = \mathbf{X}(\lambda)\hat{\boldsymbol{\beta}} = \mathbf{X}(\lambda) \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' \mathbf{y}$$

## 2.4 Regresi Campuran Nonparametrik Spline *Truncated* dan Kernel

Diberikan data berpasangan  $(x_i, t_i, y_i)$  dan hubungan antara variabel prediktor  $x_i, t_i$  dan variabel respon  $y_i$  diasumsikan mengikuti model regresi nonparametrik:

$$y_i = \mu(x_i, t_i) + \varepsilon_i, \text{ untuk } i = 1, 2, \dots, n, \quad (2.15)$$

Dalam bentuk matrik,  $\boldsymbol{\mu}$  dapat ditulis dan diuraikan:

$$\boldsymbol{\mu}(x, t) = \mathbf{f}(x) + \mathbf{g}(t) \quad (2.16)$$

Untuk mendapatkan estimator campuran regresi Spline dan Kernel, pertama kurva regresi  $f(x)$  dihampiri dengan fungsi Spline *truncated* derajat  $m$  dan titik knot  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_r)'$ . Selanjutnya kurva regresi  $g(t)$  dihampiri dengan fungsi kernel. Dengan menggunakan metode MLE, Budiantara *et al.* (2015) memperoleh estimator gabungan  $\mu(x, t)$ :

$$\hat{\mu}_{\alpha, \lambda}(x, t) = \hat{\mathbf{f}}_{\alpha \lambda}(x, t) + \hat{\mathbf{g}}_{\alpha}(t) = [A(\lambda, \alpha) + D(\alpha)] \mathbf{y}$$

Dimana  $A(\lambda, \alpha)$  dan  $D(\alpha)$  adalah matriks yang masing-masing bergantung pada titik knot  $\lambda$  dan bandwidth  $\alpha$ .

Titik-titik knot merupakan pengontrol keseimbangan antara kemulusan kurva dan kesesuaian kurva terhadap data. Pemilihan titik-titik knot dapat dilakukan salah satunya dengan metode Generalized Cross Validation atau GCV (Eubank, 1988).

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(n^{-1} \text{tr}(\mathbf{I} - (A(\lambda, \alpha) + D(\alpha)))\right)^2}, \quad (2.17)$$

dimana  $MSE(\lambda)$  pada persamaan (2.15) adalah:

$$MSE(\lambda) = n^{-1} \sum_{i=1}^n (y_i - f(x_i))^2$$

Pemilihan titik-titik knot optimal dilakukan dengan melihat nilai GCV yang terkecil.

## 2.5 Kriteria Kebaikan Model

Nilai koefisien determinasi  $R^2$  (coefficient of determination/R-squared), merupakan salah satu kriteria kebaikan model. Nilai  $R^2$  yang tinggi adalah kriteria nilai model yang baik untuk meramalkan data. Nilai  $R^2$  didefinisikan oleh

$$R^2 = \frac{SSR}{SST}$$

dengan SSR adalah Sum of Square Regression dan SST adalah Sum of Square Total, dimana:

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

## 2.6 Tingkat Pengangguran Terbuka

Konsep definisi ketenagakerjaan yang digunakan Badan Pusat Statistik (BPS) merujuk pada rekomendasi ILO sebagaimana tercantum dalam buku *“Surveys of Economically Active Population, Employment, Unemployment, and Under Method”*, ILO 1992. Hal ini dimaksudkan, terutama agar data ketenagakerjaan yang dihasilkan dari berbagai survey di Indonesia dapat dibandingkan secara internasional, tanpa mengesampingkan kondisi ketenagakerjaan spesifik Indonesia. Menurut Badan Pusat Statistik yang dimaksud dengan angkatan kerja adalah penduduk usia 15 tahun ke atas yang bekerja atau sementara tidak bekerja, dan yang sedang mencari pekerjaan. Sementara itu yang termasuk bukan angkatan kerja adalah penduduk usia kerja (15 tahun dan lebih) yang masih sekolah, mengurus rumah tangga, atau melaksanakan kegiatan lainnya.

Pengangguran dibedakan menjadi tiga yaitu pengangguran terbuka, pengangguran terselubung, dan setengah menganggur. Pengangguran Terbuka adalah penduduk yang mencari pekerjaan, yang mempersiapkan usaha, yang tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan, yang sudah mempunyai pekerjaan tetapi belum mulai bekerja dari sejumlah angkatan kerja yang ada. Pengangguran terselubung adalah seorang yang bekerja tetapi penghasilan yang diperoleh tidak mencukupi kebutuhan hidupnya. Sementara yang dimaksud dengan setengah pengangguran adalah mereka yang bekerja di bawah jam kerja normal (kurang dari 35 jam seminggu), dan masih mencari pekerjaan atau masih bersedia menerima pekerjaan (dahulu disebut setengah pengangguran terpaksa) (BPS, 2012).

Tingkat Pengangguran Terbuka adalah perbandingan antara jumlah pencari kerja dengan jumlah angkatan kerja. TPT dihitung berdasarkan rumus berikut:

$$TPT = \frac{\text{jumlah pencari kerja}}{\text{jumlah angkatan kerja}} \times 100\% \quad (2.18)$$

Besarnya angka pengangguran terbuka mempunyai implikasi sosial yang luas karena mereka yang tidak bekerja tidak mempunyai pendapatan. Semakin tinggi angka pengangguran terbuka maka semakin besar potensi kerawanan sosial



yang ditimbulkannya, contohnya kriminalitas. Sebaliknya semakin rendah angka pengangguran terbuka maka semakin stabil kondisi sosial dalam masyarakat. Sangatlah tepat jika pemerintah seringkali menjadikan indikator ini sebagai tolok ukur keberhasilan pembangunan.

International Labour Organization (ILO) mendefinisikan beberapa peubah yang berpengaruh terhadap TPT, diantaranya kependudukan, pendidikan, upah tenaga kerja, PDRB, banyaknya pekerjaan di sektor formal dan informal, infrastruktur, serta sarana dan prasarana yang tersedia di suatu wilayah. Indikator TPT ini menjadi acuan pemerintah untuk dibukanya lapangan kerja baru (BPS, 2007).

Salah satu indikator yang paling penting dalam kependudukan adalah kepadatan penduduk. Jumlah penduduk yang terlalu banyak atau kepadatan penduduk yang terlalu tinggi akan menjadi penghambat pembangunan ekonomi di negara berkembang. Pendapatan per kapita yang rendah dan tingkat pembentukan modal yang rendah semakin sulit bagi negara berkembang untuk menopang ledakan jumlah penduduk. Sekalipun output meningkat sebagai hasil teknologi yang lebih baik dan pembentukan modal, peningkatan ini akan ditelan oleh jumlah penduduk yang terlalu banyak. Alhasil, tidak ada perbaikan dalam laju pertumbuhan nyata perekonomian, sehingga berdampak pada banyaknya pengangguran. Meningkatnya jumlah penduduk yang tidak disertai dengan jumlah lapangan kerja yang memadai akan meningkatkan jumlah pengangguran. Ketidakseimbangan antara peningkatan jumlah penduduk dengan jumlah lapangan pekerjaan yang tersedia dapat menimbulkan pengangguran (Bellante, 1990)

Menurut Sukirno (2000) pertumbuhan ekonomi berarti perkembangan kegiatan dalam perekonomian yang menyebabkan barang dan jasa yang diproduksi dalam masyarakat bertambah dan akan meningkatkan kemakmuran masyarakat. Laju pertumbuhan ekonomi juga mempengaruhi Tingkat Pengangguran Terbuka. Pertumbuhan ekonomi biasanya diikuti oleh terciptanya lapangan pekerjaan yang baru. Ketika ekonomi bertumbuh, berarti terdapat pertumbuhan produksi barang dan jasa. Ketika hal ini terjadi maka kebutuhan akan tenaga kerja untuk memproduksi barang dan jasa pun akan tumbuh. Dengan kata lain, jika pertumbuhan ekonomi meningkat maka akan mengurangi

pengangguran. Pertumbuhan ekonomi suatu Negara yang terus menerus menunjukkan peningkatan menggambarkan perekonomian negara tersebut berkembang dengan baik. Pernyataan tersebut sesuai dengan Hukum Okun (Mankiw, 2006) yang menyatakan relasi negatif antara kedua variabel tersebut.

Salah satu faktor yang juga mempengaruhi TPT yaitu pendidikan. Pendidikan sebagai sarana untuk memperoleh wawasan, ilmu pengetahuan dan keterampilan agar peluang kerja lebih terbuka dan upah yang didapat juga lebih tinggi. Masyarakat yang berpendidikan tinggi akan mempunyai keterampilan dan keahlian, sehingga dapat meningkatkan produktivitasnya. Hal ini sejalan dengan (Idris, 2007) yang menyatakan pendidikan tersebut termasuk kedalam salah satu investasi pada bidang sumber daya manusia, yang mana investasi tersebut dinamakan dengan Human Capital (teori modal manusia). Investasi pendidikan merupakan kegiatan yang dapat dinilai stock manusia, dimana nilai stock manusia setelah mengikuti pendidikan dengan berbagai jenis dan bentuk pendidikan diharapkan dapat meningkatkan berbagai bentuk nilai berupa peningkatan penghasilan individu, peningkatan produktivitas kerja, dan peningkatan nilai rasional (social benefit) individu. Salah satu variabel indikator pendidikan yang biasa digunakan dalam penelitian adalah rata-rata lama sekolah. Rata-rata lama sekolah adalah rata-rata jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun ke atas untuk menempuh semua jenis pendidikan formal yang pernah dijalannya (BPS, 2012)

Variabel yang juga berpengaruh terhadap TPT adalah upah minimum. Menurut Kusnani (1998), upah minimum adalah sebuah kontroversi, upah minimum diperlukan untuk memenuhi kebutuhan pekerja agar sampai pada tingkat pendapatan "living wage", yang berarti bahwa orang yang bekerja akan mendapatkan pendapatan yang layak untuk hidupnya. Upah minimum dapat mencegah pekerja dalam pasar monopsoni dari eksploitasi tenaga kerja terutama yang low skilled. Upah minimum dapat meningkatkan produktivitas tenaga kerja dan mengurangi konsekuensi pengangguran seperti yang diperkirakan teori ekonomi konvensional. Pada realitanya menurut Alghofari (2010) setiap kenaikan tingkat upah akan diikuti oleh turunnya tenaga kerja yang diminta, yang berarti akan menyebabkan bertambahnya pengangguran. Demikian pula sebaliknya

dengan turunnya tingkat upah maka akan diikuti oleh meningkatnya penyerapan tenaga kerja, sehingga dapat dikatakan bahwa jumlah tenaga kerja yang terserap mempunyai hubungan timbal balik dengan tingkat upah. Upah mempunyai pengaruh terhadap jumlah angkatan kerja yang bekerja. Jika semakin tinggi tingkat upah yang ditetapkan, maka berpengaruh pada meningkatnya biaya produksi, akibatnya untuk melakukan efisiensi, perusahaan terpaksa melakukan pengurangan tenaga kerja, yang berakibat pada tingginya pengangguran.

Variabel lain yang mempengaruhi TPT adalah banyaknya pekerjaan di sektor formal dan informal, infrastruktur, serta sarana dan prasarana yang tersedia di suatu wilayah. Indikator yang mewakili variabel banyaknya pekerjaan di sektor formal dan informal adalah persentase penduduk usia kerja bekerja di sektor pertanian dan persentase penduduk usia kerja bekerja di sektor informal. Indikator yang mewakili variabel infrastruktur adalah panjang jalan Kabupaten/Kota dengan kondisi rusak. Sedangkan Sarana dan Prasarana diwakili oleh variabel Persentase unit usaha industri kecil menengah per jumlah penduduk usia kerja dan persentase unit usaha industri besar per jumlah penduduk usia kerja (BPS, 2007).

## **BAB 3**

### **METODE PENELITIAN**

#### **3.1 Sumber Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang yang diperoleh dari BPS yaitu: (1) data Indikator Kesejahteraan Rakyat Provinsi Jawa Barat tahun 2013 yang merupakan hasil Survei Sosial dan Ekonomi Nasional (Susenas) 2013, (2) data keadaan Angkatan Kerja di Provinsi Jawa Barat dan 2014 yang merupakan hasil Survei Angkatan Kerja Nasional (Sakernas) 2014, dan (3) data Jawa Barat dalam Angka 2015.

#### **3.2 Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon yaitu Tingkat Pengangguran Terbuka Kabupaten/Kota di Jawa Barat, dan variabel prediktor yang terdiri dari:

**Tabel 3.1** Variabel Penelitian

<b>Variabel</b>	<b>Keterangan</b>
$y$	Tingkat Pengangguran Terbuka
$x_1$	Kepadatan Penduduk
$x_2$	Laju Pertumbuhan Ekonomi
$x_3$	Rata-rata Lama Sekolah
$x_4$	Upah Minimum
$x_5$	Persentase Penduduk Usia Kerja Bekerja di Sektor Pertanian
$x_6$	Panjang Jalan Kabupaten/Kota dengan Kondisi Rusak
$x_7$	Persentase Unit Usaha Industri Kecil Menengah per Jumlah Penduduk Usia Kerja

Definisi operasional dari variabel prediktor yang diduga berpengaruh terhadap Tingkat Pengangguran Terbuka di Jawa Barat adalah sebagai berikut:

a. Tingkat Pengangguran Terbuka (TPT)

Persentase perbandingan antara pencari kerja dengan jumlah angkatan kerja.

b. Kepadatan Penduduk

Jumlah penduduk di suatu daerah dibagi dengan luas daratan daerah tersebut, biasanya dinyatakan sebagai penduduk per km<sup>2</sup>.

c. Laju Pertumbuhan Ekonomi

Indikator yang menunjukkan sejauh mana aktifitas perekonomian dapat menghasilkan tambahan pendapatan atau kesejahteraan masyarakat pada periode tertentu

d. Rata-rata Lama Sekolah

Jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun ke atas untuk menempuh semua jenis pendidikan formal yang pernah dijalaninya dibagi dengan banyaknya penduduk yang berusia 15 tahun ke atas.

e. Upah Minimum

Standar minimum upah yang dikeluarkan dari kebijakan pemerintah kabupaten/kota.

f. Persentase Penduduk Usia Kerja Bekerja di Sektor Pertanian

Persentase banyaknya penduduk usia kerja bekerja di bidang pertanian dibagi dengan banyaknya penduduk usia kerja.

g. Persentase Panjang Jalan Kabupaten/Kota dengan Kondisi Rusak

Persentase panjang jalan baik dengan kondisi rusak biasa maupun rusak parah dibagi dengan total panjang jalan yang berada di kabupaten/kota tersebut

h. Persentase Unit Usaha Industri per Jumlah Penduduk Usia Kerja.

Persentase banyaknya unit usaha, baik usaha kecil, menengah, maupun besar dengan banyaknya penduduk yang berusia minimal lima belas tahun.

Struktur data penelitian disajikan pada Tabel 3.2. Variabel  $y$  merupakan variabel respon dan variabel  $x_1, x_2, x_3, x_4$ , dan  $x_5$  merupakan variabel prediktor.

**Tabel 3.2** Struktur Data Penelitian

Kabupaten/Kota	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
1	$y_1$	$x_{1.1}$	$x_{2.1}$	$x_{3.1}$	$x_{4.1}$	$x_{5.1}$	$x_{6.1}$	$x_{7.1}$
2	$y_2$	$x_{1.2}$	$x_{2.2}$	$x_{3.2}$	$x_{4.2}$	$x_{5.2}$	$x_{6.2}$	$x_{7.2}$
3	$y_3$	$x_{1.3}$	$x_{2.3}$	$x_{3.3}$	$x_{4.3}$	$x_{5.3}$	$x_{6.3}$	$x_{7.3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
26	$y_{26}$	$x_{1.26}$	$x_{2.26}$	$x_{3.26}$	$x_{4.26}$	$x_{5.26}$	$x_{6.26}$	$x_{7.26}$

### 3.3 Langkah-langkah Penelitian

Untuk menyelesaikan tujuan penelitian, dibuat langkah-langkah yang disusun mengikuti tahapan-tahapan berikut.

#### Tahap I. Persiapan penelitian.

Pada tahap ini dipersiapkan literatur pendukung penelitian ini kemudian melakukan studi pustaka serta mengkajinya.

#### Tahap II. Estimator Campuran Kernel dan Spline Dalam Regresi Nonparametrik Mutivariabel.

Untuk mendapatkan estimator Campuran Kernel dan Spline Dalam Regresi Nonparametrik Mutivariabel, dilakukan langkah-langkah sebagai berikut:

- 1) Mendefinisikan model regresi nonparametrik multivariabel prediktor mengikuti model aditif:

$$y_i = \sum_{s=1}^p f_s(x_{si}) + \sum_{k=1}^q g_k(t_{ki}) + \varepsilon_i, i = 1, 2, \dots, n \quad (3.1)$$

- 2) Menghampiri kurva regresi  $g_k(t_{ki})$ ,  $k = 1, 2, \dots, q$  dengan menggunakan estimator Kernel.
- 3) Menghampiri kurva regresi  $f_s(x_{si})$ ,  $s = 1, 2, \dots, p$  dengan fungsi Spline *Truncated*
- 4) Membangun komponen *Goodness of Fit*:

$$\left( \mathbf{y} - \sum_{s=1}^p \mathbf{f}_s(\boldsymbol{\lambda}, \boldsymbol{\alpha}) - \sum_{k=1}^q \hat{\mathbf{g}}_k(\boldsymbol{\alpha}) \right)' \left( \mathbf{y} - \sum_{s=1}^p \mathbf{f}_s(\boldsymbol{\lambda}, \boldsymbol{\alpha}) - \sum_{k=1}^q \hat{\mathbf{g}}_k(\boldsymbol{\alpha}) \right)$$

5) Menyelesaikan optimasi  $LS$ :

$$\underset{\beta}{Min} \left\{ \left( \mathbf{y} - \sum_{s=1}^p \mathbf{f}_s(\boldsymbol{\lambda}, \alpha) - \sum_{k=1}^q \hat{\mathbf{g}}_k(\alpha) \right)' \left( \mathbf{y} - \sum_{s=1}^p \mathbf{f}_s(\boldsymbol{\lambda}, \alpha) - \sum_{k=1}^q \hat{\mathbf{g}}_k(\alpha) \right) \right\}$$

dengan  $\boldsymbol{\lambda}$  merupakan titik knot dan  $\alpha$  parameter *bandwidth*.

6) Berdasarkan 5) diharapkan diperoleh estimator untuk Komponen Kernel, Komponen Spline dan Estimator Campuran Kernel dan Spline, berturut-turut diberikan oleh:

$\hat{\mathbf{f}}_s(\boldsymbol{\lambda}, \alpha)$ ,  $s = 1, 2, \dots, p$  (Estimator Komponen Spline),

$\hat{\mathbf{g}}_k(\alpha)$ ,  $k = 1, 2, \dots, q$  (Estimator Komponen Kernel), dan

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\lambda}, \alpha) = \sum_{s=1}^p \hat{\mathbf{f}}_s(\boldsymbol{\lambda}, \alpha) + \sum_{k=1}^q \hat{\mathbf{g}}_k(\alpha) \text{ (Estimator Campuran Kernel dan Spline)}$$

**Tahap III.** Menyelidiki sifat-sifat estimator Campuran Kernel dan Spline dalam regresi nonparametrik multivariabel.

Untuk menyelesaikan Tahap III ini, dilakukan langkah-langkah sebagai berikut:

1) Menyelidiki sifat estimator komponen Spline, apakah mempunyai sifat tidak

bias atau bias untuk kurva  $\sum_{s=1}^p \mathbf{f}_s(x_s)$ , dengan menyelidiki nilai Ekspektasi:

$$\sum_{s=1}^p E[\hat{\mathbf{f}}_s(\boldsymbol{\lambda}, \alpha)]$$

2) Menyelidiki sifat estimator komponen kernel, apakah mempunyai sifat tidak

bias atau bias untuk kurva  $\sum_{k=1}^q \mathbf{g}_k(t_k)$ , dengan menyelidiki nilai Ekspektasi:

$$\sum_{k=1}^q E[\hat{\mathbf{g}}_k(\alpha)]$$

3) Menyelidiki sifat estimator Campuran Kernel dan Spline, apakah mempunyai sifat tidak bias atau bias untuk kurva  $\boldsymbol{\mu}(x, t)$ , dengan menyelidiki nilai Ekspektasi:

$$E[\hat{\mu}(\lambda, \alpha)] = \sum_{s=1}^p E[\hat{\mathbf{f}}_s(\lambda, \alpha)] + \sum_{k=1}^q E[\hat{\mathbf{g}}_k(\alpha)]$$

**Tahap IV.** Memodelkan tingkat pengangguran terbuka di Jawa Barat dan Banten menggunakan estimator campuran Kernel dan Spline dalam regresi nonparametrik multivariabel, dengan langkah-langkah sebagai berikut:

- 1) Membuat plot secara parsial antara variabel respon dengan masing-masing variabel prediktor untuk menentukan model awal.
- 2) Menentukan variabel prediktor-prediktor yang dihipotesiskan dengan model Spline dan menentukan prediktor-prediktor yang dihipotesiskan dengan Kernel melalui plot yang telah diperoleh.
- 3) Memodelkan data dengan menggunakan estimator campuran Kernel dan Spline
- 4) Memilih titik knot dan parameter *bandwidth* optimal dengan metode GCV.
- 5) Membuat kesimpulan dan peramalan dengan model yang diperoleh.
- 6) Membuat rekomendasi kebijakan yang dapat digunakan para stakeholder dalam mengambil keputusan.



*(Halaman ini sengaja dikosongkan)*

## BAB 4

### HASIL DAN PEMBAHASAN

Pada bab ini dilakukan pembahasan berdasarkan tujuan penelitian. Penulis akan membahas estimasi parameter pada gabungan regresi nonparametrik spline *truncated* dan kernel dengan menggunakan optimasi *Least Square* (LS). Kemudian akan diselidiki sifat-sifat estimator yang diperoleh. Setelah itu, estimator tersebut akan diterapkan pada permasalahan nyata yaitu kasus Tingkat Pengangguran Terbuka provinsi Jawa Barat pada tahun 2013.

#### 4. 1. Estimasi Model Campuran Regresi Nonparametrik Spline dan Kernel

Diberikan data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, y_i)$  dan hubungan antara variabel prediktor  $x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}$  dan variabel respon  $y_i$  diasumsikan mengikuti model regresi nonparametrik:

$$y_i = \mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (4.1)$$

Bentuk kurva regresi  $\mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi})$  diasumsikan tidak diketahui dan hanya diasumsikan *smooth* yang artinya kontinu dan differentiable. Error random  $\varepsilon_i$  berdistribusi Normal dengan mean nol dan varians  $\sigma^2$ . Kurva regresi  $\mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi})$  diasumsikan bersifat aditif, sehingga dapat ditulis:

$$\mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}) = f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_p(x_{pi}) + g_1(t_{1i}) + g_2(t_{2i}) + \dots + g_q(t_{qi}), \quad (4.2)$$

dapat ditulis menjadi:

$$\mu(x, t) = \sum_{s=1}^p f_s(x_{si}) + \sum_{k=1}^q g_k(t_{ki}), \quad (4.3)$$

dengan  $f_s(x_{si})$  dan  $g_k(t_{ki})$  merupakan fungsi yang *smooth*. Persoalan utama dalam estimator campuran kurva regresi nonparametrik adalah mendapatkan bentuk estimasi kurva regresi  $\mu(x, t)$

$$\hat{\mu}(x, t) = \sum_{s=1}^p \hat{f}_s(x, t) + \sum_{k=1}^q \hat{g}_k(t). \quad (4.4)$$

Untuk mendapatkan estimator campuran regresi Spline dan Kernel, pertama kurva regresi  $f_s(x_{si})$  dihampiri dengan fungsi Spline truncated derajat  $m_s$  dan titik knot  $\lambda_s = (\lambda_{s1}, \lambda_{s2}, \dots, \lambda_{sr})'$ . Selanjutnya kurva regresi  $g_k(t_{ki})$  dihampiri dengan fungsi Kernel. Misalkan diberikan basis untuk ruang sampel Spline  $1, x, x^2, \dots, x^{m_s}, (x - \lambda_{s1})^{m_s} I(x \geq \lambda_{s1}), \dots, (x - \lambda_{sr})^{m_s} I(x \geq \lambda_{sr})$ , dengan  $I$  merupakan fungsi indikator. Kurva regresi  $f_s(x_{si})$  dapat ditulis menjadi:

$$f_s(x_{si}) = \theta_{s0} + \theta_{s1}x_{si} + \theta_{s2}x_{si}^2 + \dots + \theta_{sm_s}x_{si}^{m_s} + \phi_{s1}(x_{si} - \lambda_{s1})_+^{m_s} + \dots + \phi_{sr}(x_{si} - \lambda_{sr})_+^{m_s}, \quad (4.5)$$

dengan  $\theta_{s0}, \theta_{s1}, \dots, \theta_{sm_s}, \phi_{s1}, \dots, \phi_{sr}$  merupakan parameter-parameter yang tidak diketahui. Untuk estimasi kurva regresi  $g_k(t_{ki})$  dapat disajikan menjadi:

$$\hat{g}_{\alpha_k}(t) = n^{-1} \sum_{i=1}^n \left( \frac{K_{\alpha_k}(t - t_i)}{n^{-1} \sum_{j=1}^n K_{\alpha_k}(t - t_j)} \right) y_i = n^{-1} \sum_{i=1}^n W_{\alpha_k i}(t) y_i, \quad (4.6)$$

dimana:

$$W_{\alpha_k i}(t) = \frac{K_{\alpha_k}(t - t_i)}{n^{-1} \sum_{j=1}^n K_{\alpha_k}(t - t_j)}, \quad K_{\alpha_k}(t - t_i) = \frac{1}{\alpha_k} K\left(\frac{(t - t_i)}{\alpha_k}\right),$$

dengan  $K$  merupakan fungsi Kernel Gaussian, dengan persamaan:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty$$

Estimator kurva regresi (4.6) sangat bergantung kepada dua hal yaitu parameter bandwidth dan fungsi Kernelnya. Berdasarkan estimator (4.5), akan dicari estimator campuran regresi Spline dan Kernel pada persamaan (4.4) menggunakan metode MLE. Untuk tujuan ini, diberikan beberapa teorema dan lemma berikut:

### Lemma 1

Jika kurva regresi  $f_s(x_{si})$  diberikan pada persamaan (4.5), maka:

$$\mathbf{f} = \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\Phi}$$

dimana vektor  $\mathbf{f}$ ,  $\boldsymbol{\theta}$ , dan  $\boldsymbol{\Phi}$  berturut-turut adalah:

$$\mathbf{f} = (\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_p)' , \quad \mathbf{f}_i = (f_i(x_{i1}) \quad f_i(x_{i2}) \quad \cdots \quad f_i(x_{in}))' , \quad i = 1, 2, \dots, p \quad (4.7)$$

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad \cdots \quad \boldsymbol{\theta}_p)' , \quad \boldsymbol{\theta}_j = (\theta_{j0} \quad \theta_{j1} \quad \cdots \quad \theta_{jm_j})' , \quad j = 1, 2, 3, \dots, p \quad (4.8)$$

$$\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \cdots \quad \boldsymbol{\Phi}_p)' , \quad \boldsymbol{\Phi}_l = (\phi_{l0} \quad \phi_{l1} \quad \cdots \quad \phi_{lm_l})' , \quad l = 1, 2, 3, \dots, p \quad (4.9)$$

dan matriks  $X_1$  dan  $X_2(\tilde{\lambda})$  berturut-turut adalah :

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^{m_1} & 1 & \cdots & x_{21}^{m_2} & 1 & \cdots & x_{p1}^{m_p} \\ 1 & x_{12} & \cdots & x_{12}^{m_1} & 1 & \cdots & x_{22}^{m_2} & 1 & \cdots & x_{p2}^{m_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^{m_1} & 1 & \cdots & x_{2n}^{m_2} & 1 & \cdots & x_{pn}^{m_p} \end{bmatrix} , \text{ dan} \quad (4.10)$$

$$\mathbf{X}_2(\boldsymbol{\lambda}) = \begin{bmatrix} (x_{11} - \lambda_{11})_+^{m_1} & \cdots & (x_{21} - \lambda_{21})_+^{m_2} & \cdots & (x_{p1} - \lambda_{pr_p})_+^{m_p} \\ (x_{12} - \lambda_{11})_+^{m_1} & \cdots & (x_{22} - \lambda_{21})_+^{m_2} & \cdots & (x_{p2} - \lambda_{pr_p})_+^{m_p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (x_{1n} - \lambda_{11})_+^{m_1} & \cdots & (x_{2n} - \lambda_{21})_+^{m_2} & \cdots & (x_{pn} - \lambda_{pr_p})_+^{m_p} \end{bmatrix} \quad (4.11)$$

### Bukti:

Dari persamaan (4.5)

$$f_s(x_{si}) = \theta_{s0} + \theta_{s1}x_{si} + \theta_{s2}x_{si}^2 + \dots + \theta_{sm_s}x_{si}^{m_s} + \phi_{s1}(x_{si} - \lambda_{s1})_+^{m_s} + \dots + \phi_{sr_s}(x_{si} - \lambda_{sr_s})_+^{m_s}$$

Untuk  $s = 1$ , diperoleh:

$$\begin{aligned} f_1(x_{1i}) &= \theta_{10} + \theta_{11}x_{1i} + \theta_{12}x_{1i}^2 + \dots + \theta_{1m_1}x_{1i}^{m_1} + \phi_{11}(x_{1i} - \lambda_{11})_+^{m_1} + \dots + \phi_{1r_1}(x_{1i} - \lambda_{1r_1})_+^{m_1} \\ f_1(x_{11}) &= \theta_{10} + \theta_{11}x_{11} + \theta_{12}x_{11}^2 + \dots + \theta_{1m_1}x_{11}^{m_1} + \phi_{11}(x_{11} - \lambda_{11})_+^{m_1} + \dots + \phi_{1r_1}(x_{11} - \lambda_{1r_1})_+^{m_1} \\ f_1(x_{12}) &= \theta_{10} + \theta_{11}x_{12} + \theta_{12}x_{12}^2 + \dots + \theta_{1m_1}x_{12}^{m_1} + \phi_{11}(x_{12} - \lambda_{11})_+^{m_1} + \dots + \phi_{1r_1}(x_{12} - \lambda_{1r_1})_+^{m_1} \\ &\vdots \\ f_1(x_{1n}) &= \theta_{10} + \theta_{11}x_{1n} + \theta_{12}x_{1n}^2 + \dots + \theta_{1m_1}x_{1n}^{m_1} + \phi_{11}(x_{1n} - \lambda_{11})_+^{m_1} + \dots + \phi_{1r_1}(x_{1n} - \lambda_{1r_1})_+^{m_1} \end{aligned}$$

Dapat diubah dalam bentuk matiks menjadi:

$$\begin{pmatrix} f_1(x_{11}) \\ f_1(x_{12}) \\ \vdots \\ f_1(x_{1n}) \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{11}^{m_1} \\ 1 & x_{12} & \cdots & x_{12}^{m_1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^{m_1} \end{pmatrix} \begin{pmatrix} \theta_{10} \\ \theta_{11} \\ \vdots \\ \theta_{1m_1} \end{pmatrix} + \begin{pmatrix} (x_{11} - \lambda_{11})_+^{m_1} & \cdots & (x_{11} - \lambda_{1r_1})_+^{m_1} \\ (x_{12} - \lambda_{11})_+^{m_1} & \cdots & (x_{12} - \lambda_{1r_1})_+^{m_1} \\ \vdots & \vdots & \vdots \\ (x_{1n} - \lambda_{11})_+^{m_1} & \cdots & (x_{1n} - \lambda_{1r_1})_+^{m_1} \end{pmatrix} \begin{pmatrix} \phi_{11} \\ \phi_{12} \\ \vdots \\ \phi_{1r_1} \end{pmatrix}$$

$$\mathbf{f}_1 = \mathbf{X}_{11}\boldsymbol{\theta}_1 + \mathbf{X}_{12}(\boldsymbol{\lambda}_1)\boldsymbol{\Phi}_1$$

Untuk  $s = 2, 3, \dots, n$  dilakukan hal yang sama, sehingga diperoleh:

$s = 2, \mathbf{f}_2 = \mathbf{X}_{21}\boldsymbol{\theta}_2 + \mathbf{X}_{22}(\boldsymbol{\lambda}_2)\boldsymbol{\Phi}_2$ ,  $s = 3, \mathbf{f}_3 = \mathbf{X}_{31}\boldsymbol{\theta}_3 + \mathbf{X}_{32}(\boldsymbol{\lambda}_3)\boldsymbol{\Phi}_3$ , sampai dengan  $s = p, \mathbf{f}_p = \mathbf{X}_{p1}\boldsymbol{\theta}_p + \mathbf{X}_{p2}(\boldsymbol{\lambda}_p)\boldsymbol{\Phi}_p$ . Sehingga diperoleh:

$$\begin{aligned} \mathbf{f} &= \sum_{s=1}^p \mathbf{f}_s \\ &= \mathbf{f}_1 + \mathbf{f}_2 + \dots + \mathbf{f}_p \\ &= \mathbf{X}_{11}\boldsymbol{\theta}_1 + \mathbf{X}_{12}(\boldsymbol{\lambda}_1)\boldsymbol{\Phi}_1 + \mathbf{X}_{21}\boldsymbol{\theta}_2 + \mathbf{X}_{22}(\boldsymbol{\lambda}_2)\boldsymbol{\Phi}_2 + \dots + \mathbf{X}_{p1}\boldsymbol{\theta}_p + \mathbf{X}_{p2}(\boldsymbol{\lambda}_p)\boldsymbol{\Phi}_p \\ &= \mathbf{X}_{11}\boldsymbol{\theta}_1 + \mathbf{X}_{21}\boldsymbol{\theta}_2 + \dots + \mathbf{X}_{p1}\boldsymbol{\theta}_p + \mathbf{X}_{12}(\boldsymbol{\lambda}_1)\boldsymbol{\Phi}_1 + \mathbf{X}_{22}(\boldsymbol{\lambda}_2)\boldsymbol{\Phi}_2 + \dots + \mathbf{X}_{p2}(\boldsymbol{\lambda}_p)\boldsymbol{\Phi}_p \\ &= \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{21} & \cdots & \mathbf{X}_{p1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_p \end{pmatrix} + \begin{pmatrix} \mathbf{X}_{12}(\boldsymbol{\lambda}_1) & \mathbf{X}_{22}(\boldsymbol{\lambda}_2) & \cdots & \mathbf{X}_{p2}(\boldsymbol{\lambda}_p) \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \\ \vdots \\ \boldsymbol{\Phi}_p \end{pmatrix} \\ &= \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2(\boldsymbol{\lambda})\boldsymbol{\Phi} \end{aligned}$$

dengan vektor  $\boldsymbol{\theta}, \boldsymbol{\Phi}$ , dan matrik  $\mathbf{X}_1, \mathbf{X}_2(\boldsymbol{\lambda})$  berturut-turut diberikan oleh persamaan (4.8), (4.9), (4.10), dan (4.11).

## Lemma 2

Jika diberikan model regresi (4.1) dan estimator untuk regresi kernal diberikan oleh persamaan (4.6) maka jumlah kuadrat error diberikan oleh:

$$\|\boldsymbol{\varepsilon}\|^2 = \|[\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\boldsymbol{\lambda})\|^2,$$

dengan  $\|\boldsymbol{\varepsilon}\|^2$  merupakan panjang vektor  $\boldsymbol{\varepsilon}$ ,  $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)'$ ,

$$\mathbf{X}(\boldsymbol{\lambda}) = (\mathbf{X}_1 \ \mathbf{X}_2(\boldsymbol{\lambda})), \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\Phi} \end{pmatrix}, \text{ dan}$$

$$\mathbf{M}(\alpha) = \begin{pmatrix} \sum_{k=1}^q W_{\alpha_k 1}(t_1) & \sum_{k=1}^q W_{\alpha_k 2}(t_1) & \cdots & \sum_{k=1}^q W_{\alpha_k n}(t_1) \\ \sum_{k=1}^q W_{\alpha_k 1}(t_2) & \sum_{k=1}^q W_{\alpha_k 2}(t_2) & \cdots & \sum_{k=1}^q W_{\alpha_k n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^q W_{\alpha_k 1}(t_n) & \sum_{k=1}^q W_{\alpha_k 2}(t_n) & \cdots & \sum_{k=1}^q W_{\alpha_k n}(t_n) \end{pmatrix} \quad (4.12)$$

**Bukti:**

Diberikan persamaan (6)  $\hat{g}_{\alpha_k}(t_i) = n^{-1} \sum_{j=1}^n W_{\alpha_k}(t_i) y_j$ ,  $i = 1, 2, \dots$ , untuk masing-

masing  $k = 1, 2, \dots, q$ , sehingga untuk  $k=1$ , diperoleh  $\hat{g}_{\alpha_1}(t_i) = n^{-1} \sum_{j=1}^n W_{\alpha_1 j}(t_i) y_j$ .

Karena berlaku untuk  $i = 1, 2, \dots, n$ , maka:

$$\begin{aligned} \hat{g}_{\alpha_1}(t_1) &= W_{\alpha_1 1}(t_1) y_1 + W_{\alpha_1 2}(t_1) y_2 + \dots + W_{\alpha_1 n}(t_1) y_n \\ \hat{g}_{\alpha_1}(t_2) &= W_{\alpha_1 1}(t_2) y_1 + W_{\alpha_1 2}(t_2) y_2 + \dots + W_{\alpha_1 n}(t_2) y_n \\ &\vdots \\ \hat{g}_{\alpha_1}(t_n) &= W_{\alpha_1 1}(t_n) y_1 + W_{\alpha_1 2}(t_n) y_2 + \dots + W_{\alpha_1 n}(t_n) y_n \end{aligned}$$

dalam bentuk matriks, akan diperoleh :

$$\begin{pmatrix} \hat{g}_{\alpha_1}(t_1) \\ \hat{g}_{\alpha_1}(t_2) \\ \vdots \\ \hat{g}_{\alpha_1}(t_n) \end{pmatrix} = \begin{pmatrix} W_{\alpha_1 1}(t_1) & W_{\alpha_1 2}(t_1) & \cdots & W_{\alpha_1 n}(t_1) \\ W_{\alpha_1 1}(t_2) & W_{\alpha_1 2}(t_2) & \cdots & W_{\alpha_1 n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ W_{\alpha_1 1}(t_n) & W_{\alpha_1 2}(t_n) & \cdots & W_{\alpha_1 n}(t_n) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\hat{\mathbf{g}}_1 = \mathbf{M}(\alpha_1) \mathbf{y} \quad (4.13)$$

Untuk  $k = 2, 3, \dots, q$  diberikan perlakuan yang sama, dan akan diperoleh

untuk  $k = 2$ ,  $\hat{\mathbf{g}}_2 = \mathbf{M}(\alpha_2) \mathbf{y}$ ,  $k = 3$ ,  $\hat{\mathbf{g}}_3 = \mathbf{M}(\alpha_3) \mathbf{y}$  sampai  $k = q$ ,  $\hat{\mathbf{g}}_q = \mathbf{M}(\alpha_q) \mathbf{y}$ , diperoleh:

$$\begin{aligned} \sum_{k=1}^q \hat{\mathbf{g}}_k &= \hat{\mathbf{g}}_1 + \hat{\mathbf{g}}_2 + \dots + \hat{\mathbf{g}}_q \\ &= \mathbf{M}(\alpha_1) \mathbf{y} + \mathbf{M}(\alpha_2) \mathbf{y} + \dots + \mathbf{M}(\alpha_q) \mathbf{y} \\ &= (\mathbf{M}(\alpha_1) + \mathbf{M}(\alpha_2) + \dots + \mathbf{M}(\alpha_q)) \mathbf{y} \\ &= \mathbf{M}(\alpha) \mathbf{y} \end{aligned}$$

dimana:

$$\begin{aligned}
\mathbf{M}(\alpha) &= \mathbf{M}(\alpha_1) + \mathbf{M}(\alpha_2) + \dots + \mathbf{M}(\alpha_q) \\
&= \begin{pmatrix} W_{\alpha_1 1}(t_1) & W_{\alpha_1 2}(t_1) & \dots & W_{\alpha_1 n}(t_1) \\ W_{\alpha_1 1}(t_2) & W_{\alpha_1 2}(t_2) & \dots & W_{\alpha_1 n}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ W_{\alpha_1 1}(t_n) & W_{\alpha_1 2}(t_n) & \dots & W_{\alpha_1 n}(t_n) \end{pmatrix} + \begin{pmatrix} W_{\alpha_2 1}(t_1) & W_{\alpha_2 2}(t_1) & \dots & W_{\alpha_2 n}(t_1) \\ W_{\alpha_2 1}(t_2) & W_{\alpha_2 2}(t_2) & \dots & W_{\alpha_2 n}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ W_{\alpha_2 1}(t_n) & W_{\alpha_2 2}(t_n) & \dots & W_{\alpha_2 n}(t_n) \end{pmatrix} + \\
&\dots + \begin{pmatrix} W_{\alpha_q 1}(t_1) & W_{\alpha_q 2}(t_1) & \dots & W_{\alpha_q n}(t_1) \\ W_{\alpha_q 1}(t_2) & W_{\alpha_q 2}(t_2) & \dots & W_{\alpha_q n}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ W_{\alpha_q 1}(t_n) & W_{\alpha_q 2}(t_n) & \dots & W_{\alpha_q n}(t_n) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{k=1}^q W_{\alpha_k 1}(t_1) & \sum_{k=1}^q W_{\alpha_k 2}(t_1) & \dots & \sum_{k=1}^q W_{\alpha_k n}(t_1) \\ \sum_{k=1}^q W_{\alpha_k 1}(t_2) & \sum_{k=1}^q W_{\alpha_k 2}(t_2) & \dots & \sum_{k=1}^q W_{\alpha_k n}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{k=1}^q W_{\alpha_k 1}(t_n) & \sum_{k=1}^q W_{\alpha_k 2}(t_n) & \dots & \sum_{k=1}^q W_{\alpha_k n}(t_n) \end{pmatrix}
\end{aligned}$$

Model regresi (4.1), persamaan (4.13), dan lemma 1, memberikan:

$$\begin{aligned}
\mathbf{y} &= \sum_{s=1}^p \mathbf{f}_s + \sum_{k=1}^q \hat{\mathbf{g}}_k + \boldsymbol{\varepsilon} \\
&= \mathbf{X}_1 \boldsymbol{\theta} + \mathbf{X}_2(\boldsymbol{\lambda}) \boldsymbol{\Phi} + \mathbf{M}(\alpha) \mathbf{y} + \boldsymbol{\varepsilon} \\
&= (\mathbf{X}_1 \quad \mathbf{X}_2(\boldsymbol{\lambda})) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\Phi} \end{pmatrix} + \mathbf{M}(\alpha) \mathbf{y} + \boldsymbol{\varepsilon} \\
&= \mathbf{X} \boldsymbol{\beta} + \mathbf{M}(\alpha) \mathbf{y} + \boldsymbol{\varepsilon}
\end{aligned} \tag{4.14}$$

Persamaan (4.11) memberikan jumlah kuadrat error:

$$\begin{aligned}
\|\boldsymbol{\varepsilon}\|^2 &= \|\mathbf{y} - \mathbf{X}(\boldsymbol{\lambda}) \boldsymbol{\beta} - \mathbf{M}(\alpha) \mathbf{y}\|^2 \\
&= \|[\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\boldsymbol{\lambda})\|^2
\end{aligned}$$

### Teorema 1

Jika jumlah kuadrat error dari model regresi nonparametric diberikan oleh Lemma 2, error model berdistribusi Normal multivariate dengan mean nol dan  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}$  dan  $L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda)$  adalah fungsi likelihood, maka estimator MLE untuk vektor parameter  $\boldsymbol{\beta}$  diperoleh dari optimasi :

$$\underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Min}} \left\{ \left\| [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\lambda) \boldsymbol{\beta} \right\|^2 \right\}$$

### Bukti:

Diberikan model regresi nonparametrik seperti model (1),  $\boldsymbol{\varepsilon}$  berdistribusi normal mutivariat dengan  $E(\boldsymbol{\varepsilon}) = 0$  dan  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}$ , maka fungsi likelihood  $L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda)$  diberikan oleh:

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) &= \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2}} \text{Exp} \left( -\frac{\varepsilon_i^2}{2\sigma^2} \right) \right) \\ &= (2\pi\sigma^2)^{-n/2} \text{Exp} \left( -\frac{1}{2\sigma^2} \|\boldsymbol{\varepsilon}\|^2 \right) \end{aligned}$$

Berdasarkan Lemma 2 diperoleh fungsi likelihood:

$$L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) = (2\pi\sigma^2)^{-n/2} \text{Exp} \left( -\frac{1}{2\sigma^2} \left\| [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\lambda) \boldsymbol{\beta} \right\|^2 \right)$$

Berdasarkan metode MLE, estimator untuk parameter  $\boldsymbol{\beta}$  diperoleh:

$$\underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ (2\pi\sigma^2)^{-n/2} \text{Exp} \left( -\frac{1}{2\sigma^2} \left\| [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\lambda) \boldsymbol{\beta} \right\|^2 \right) \right\}$$

Selanjutnya fungsi log likelihoodnya diberikan oleh:

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) &= \log L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\| [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\lambda) \boldsymbol{\beta} \right\|^2 \end{aligned}$$

$$\text{Akibatnya: } \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ L(\boldsymbol{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\| [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} - \mathbf{X}(\lambda) \boldsymbol{\beta} \right\|^2 \right\} \quad (4.15)$$

Persamaan (4.15) mencapai nilai maksimum jika komponen log likelihood



$\|[\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta}\|^2$  mempunyai nilai minimum. Akibatnya:

$$\underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Max}} \left\{ L(\boldsymbol{\beta}, \sigma^2 \mid \alpha, \lambda) \right\} = \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Min}} \left\{ \left\| [\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta} \right\|^2 \right\}$$

Selanjutnya berdasarkan Lemma 1, Lemma 2 dan Teorema 1, dapat dicari estimator campuran regresi Spline dan Kernel, yang selengkapnya diberikan dalam Teorema 2.

### **Teorema 2**

Jika diberikan model regresi (1), jumlah kuadrat error diberikan oleh Lemma 2, error model berdistribusi Normal multivariate dengan mean nol dan  $E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$ , dan estimator MLE untuk parameter  $\boldsymbol{\beta}$  diperoleh dari optimasi seperti pada Teorema 1, maka estimator MLE untuk kurva regresi campuran  $\mu(x, t)$  diberikan oleh :

$$\hat{\mu}_{\alpha\lambda}(x, t) = \hat{\mathbf{f}}_{\alpha\lambda}(x, t) + \hat{\mathbf{g}}_{\alpha}(t)$$

dimana  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{X}(\lambda)\hat{\boldsymbol{\beta}}(\lambda, \alpha)$ ,  $\hat{\mathbf{g}}_{\alpha}(t) = \mathbf{M}(\alpha)\mathbf{y}$ , dan

$$\hat{\boldsymbol{\beta}}(\lambda, \alpha) = \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha))\mathbf{y}, \text{ dengan matrik } \mathbf{X}(\lambda) \text{ dan}$$

$\mathbf{M}(\alpha)$  diberikan oleh lemma 2.

### **Bukti:**

Untuk mendapatkan estimator MLE untuk vektor parameter  $\boldsymbol{\beta}$ , teorema 2 memberikan optimasi :

$$\underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Min}} \left\{ \left\| [\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta} \right\|^2 \right\} = \underset{\substack{\boldsymbol{\beta} \in R^{i=1} \\ \sum_{i=1}^p (m_i + \eta_i + 1)}}{\text{Min}} \left\{ Q(\boldsymbol{\beta}, \sigma^2 \mid \alpha, \lambda) \right\}$$

dengan:

$$\begin{aligned} Q(\boldsymbol{\beta}, \sigma^2 \mid \alpha, \lambda) &= \left\| [\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta} \right\|^2 \\ &= ([\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta})' ([\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta}) \\ &= \left( \mathbf{y}'(\mathbf{I} - \mathbf{M}(\alpha))' - \boldsymbol{\beta}'(\mathbf{X}(\lambda))' \right) ([\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} - \mathbf{X}(\lambda)\boldsymbol{\beta}) \\ &= \left\| [\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} \right\|^2 - 2\boldsymbol{\beta}' \left( (\mathbf{X}(\lambda))' \right) [\mathbf{I} - \mathbf{M}(\alpha)]\mathbf{y} + \boldsymbol{\beta}'(\mathbf{X}(\lambda))' \mathbf{X}(\lambda)\boldsymbol{\beta} \end{aligned}$$

Selanjutnya dengan derivatif parsial:

$$\frac{\partial}{\partial \boldsymbol{\beta}} [Q(\boldsymbol{\beta}, \sigma^2 | \alpha, \boldsymbol{\lambda})] = 0$$

Memberikan persamaan normal:

$$(\mathbf{X}(\boldsymbol{\lambda}))' \mathbf{X}(\boldsymbol{\lambda}) \boldsymbol{\beta} = (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{I} - \mathbf{M}(\alpha)) \mathbf{y} \quad (4.16)$$

Estimator MLE untuk  $\boldsymbol{\beta}$  diberikan oleh:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha) = \left[ (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{X}(\boldsymbol{\lambda})) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} = \mathbf{C}(\boldsymbol{\lambda}, \alpha) \mathbf{y} \quad , \quad (4.17)$$

$$\text{dengan } \mathbf{C}(\boldsymbol{\lambda}, \alpha) \mathbf{y} = \left[ (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{X}(\boldsymbol{\lambda})) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{I} - \mathbf{M}(\alpha)) \mathbf{y} .$$

Mengingat persamaan (4.17) dan sifat invarians dari metode MLE, maka estimator untuk kurva spline  $\mathbf{f}(x, t) = \mathbf{X}(\boldsymbol{\lambda}) \boldsymbol{\beta}$  diberikan oleh  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{X}(\boldsymbol{\lambda}) \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha)$  dan telah diperoleh (4.14) estimator kernel  $\hat{\mathbf{g}}_{\alpha}(t) = \mathbf{M}(\alpha) \mathbf{y}$ . Lemma 1 memberikan kurva regresi campuran regresi spline dan kernel:  $\boldsymbol{\mu}(x, t) = \mathbf{f}(x) + \mathbf{g}(t)$ . Akibatnya estimator campuran regresi spline dan kernel  $\boldsymbol{\mu}(x, t)$  diberikan oleh:

$$\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t) = \hat{\mathbf{f}}_{\alpha, \lambda}(x, t) + \hat{\mathbf{g}}_{\alpha}(t)$$

### Lemma 3

Jika estimator-estimator  $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha)$ ,  $\hat{\mathbf{g}}_{\alpha}(t)$ ,  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t)$ , dan  $\hat{\boldsymbol{\mu}}_{\alpha\lambda}(x, t)$  diberikan dalam Teorema 2, maka:

$$\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{N}(\boldsymbol{\lambda}, \alpha) \mathbf{y}, \text{ dan } \hat{\boldsymbol{\mu}}_{\alpha\lambda}(x, t) = \mathbf{P}(\boldsymbol{\lambda}, \alpha) \mathbf{y}$$

$$\text{dengan } \mathbf{N}(\boldsymbol{\lambda}, \alpha) = \mathbf{X}(\boldsymbol{\lambda}) \left[ (\mathbf{X}(\boldsymbol{\lambda}))' \mathbf{X}(\boldsymbol{\lambda}) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{I} - \mathbf{M}(\alpha)), \text{ dan}$$

$$\mathbf{P}(\boldsymbol{\lambda}, \alpha) = \mathbf{N}(\boldsymbol{\lambda}, \alpha) + \mathbf{M}(\alpha)$$

**Bukti:**

Teorema 2 memberikan  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{X}(\lambda)\hat{\boldsymbol{\beta}}(\lambda, \alpha)$ . Persamaan (4.13) dan (4.14)

memberikan  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{X}(\lambda) \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha)) \mathbf{y}$ , dengan

matrik  $\mathbf{N}(\lambda, \alpha) = \mathbf{X}(\lambda) \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha))$ . Teorema 2, dan

persamaan (4.14) memberikan:

$$\hat{\boldsymbol{\mu}}_{\alpha,\lambda}(x, t) = \hat{\mathbf{f}}_{\alpha\lambda}(x, t) + \hat{\mathbf{g}}_{\alpha}(t) = \mathbf{N}(\lambda, \alpha) \mathbf{y} + \mathbf{M}(\alpha) \mathbf{y} = [\mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)] \mathbf{y} = \mathbf{P}(\lambda, \alpha) \mathbf{y}$$

dengan  $\mathbf{P}(\lambda, \alpha) = \mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)$ . Maka  $\hat{\mathbf{f}}_{\alpha\lambda}(x, t) = \mathbf{N}(\lambda, \alpha) \mathbf{y}$ , dan

$$\hat{\boldsymbol{\mu}}_{\alpha\lambda}(x, t) = \mathbf{P}(\lambda, \alpha) \mathbf{y} \text{ dengan } \mathbf{N}(\lambda, \alpha) = \mathbf{X}(\lambda) \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha))$$

$$\text{dan } \mathbf{P}(\lambda, \alpha) = \mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)$$

**4. 2. Sifat Estimator Campuran Regresi Spline dan Kernel**

Berikut ini akan diselediki sifat-sifat dari estimator  $\hat{\boldsymbol{\beta}}(\lambda, \alpha)$ , estimator regresi Kernel  $\hat{\mathbf{g}}_{\alpha}(t)$ , estimator regresi Spline  $\hat{\mathbf{f}}_{\alpha,\lambda}(x, t)$  dan estimator campuran regresi Spline dan Kernel  $\hat{\boldsymbol{\mu}}_{\alpha,\lambda}(x, t)$ .

Jika  $\hat{\boldsymbol{\beta}}(\lambda, \alpha)$ ,  $\hat{\mathbf{g}}_{\alpha}(t)$ ,  $\hat{\mathbf{f}}_{\alpha,\lambda}(x, t)$  dan  $\hat{\boldsymbol{\mu}}_{\alpha,\lambda}(x, t)$  adalah estimator yang diberikan dalam Teorema 2 dan lemma 3, maka :

a. Sifat estimator  $\hat{\boldsymbol{\beta}}(\lambda, \alpha)$

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}(\lambda, \alpha)] &= \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha)) E(\mathbf{y}) \\ &= \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha)) \left[ \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) \right] \\ &= \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' (\mathbf{I} - \mathbf{M}(\alpha)) \left[ \mathbf{X}(\lambda) \boldsymbol{\beta} + \sum_{k=1}^q \mathbf{g}_k(t) \right] \end{aligned}$$

$$\begin{aligned}
&= \boldsymbol{\beta} + \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' \sum_{k=1}^q \mathbf{g}_k(t) - \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} \\
&\quad (\mathbf{X}(\lambda))' \mathbf{M}(\alpha) \mathbf{X}(\lambda) \boldsymbol{\beta} - \left[ (\mathbf{X}(\lambda))' \mathbf{X}(\lambda) \right]^{-1} (\mathbf{X}(\lambda))' \mathbf{M}(\alpha) \sum_{k=1}^q \mathbf{g}_k(t) \\
&\neq \boldsymbol{\beta}
\end{aligned}$$

b. Sifat estimator  $\hat{\mathbf{g}}_{\alpha}(t)$

$$\begin{aligned}
\mathbb{E} \left[ \sum_{k=1}^q \hat{\mathbf{g}}_k(t) \right] &= \mathbb{E} [\mathbf{M}(\alpha) \mathbf{y}] \\
&= \mathbf{M}(\alpha) \mathbb{E} [\mathbf{y}] \\
&= \mathbf{M}(\alpha) \left[ \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) \right] \\
&= \mathbf{M}(\alpha) \sum_{k=1}^q \mathbf{g}_k(t) + \mathbf{M}(\alpha) \sum_{s=1}^p \mathbf{f}_s(x) \neq \sum_{k=1}^q \mathbf{g}_k(t)
\end{aligned}$$

c. Sifat estimator  $\hat{\mathbf{f}}_{\alpha, \lambda}(x, t)$

$$\begin{aligned}
\mathbb{E} \left[ \sum_{s=1}^p \hat{\mathbf{f}}_s(x) \right] &= \mathbb{E} [\mathbf{N}(\lambda, \alpha) \mathbf{y}] \\
&= \mathbf{N}(\lambda, \alpha) \left[ \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) \right] \\
&= \mathbf{N}(\lambda, \alpha) \sum_{s=1}^p \mathbf{f}_s(x) + \mathbf{N}(\lambda, \alpha) \sum_{k=1}^q \mathbf{g}_k(t) \neq \sum_{s=1}^p \mathbf{f}_s(x)
\end{aligned}$$

d. Sifat estimator  $\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t)$

$$\begin{aligned}
\mathbb{E} [\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t)] &= \mathbb{E} \left[ \sum_{s=1}^p \hat{\mathbf{f}}_s(x, t) + \sum_{k=1}^q \hat{\mathbf{g}}_k(t) \right] \\
&= \mathbb{E} [\hat{\mathbf{f}}_{\alpha \lambda}(x, t)] + \mathbb{E} [\hat{\mathbf{g}}_{\alpha}(t)] \\
&= [\mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)] \sum_{s=1}^p \mathbf{f}_s(x) + [\mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)] \sum_{k=1}^q \mathbf{g}_k(t) \\
&= [\mathbf{N}(\lambda, \alpha) + \mathbf{M}(\alpha)] \left( \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) \right) \\
&\neq \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) = \boldsymbol{\mu}(x, t)
\end{aligned}$$

Dari keempat nilai yang diperoleh, dapat dikatakan bahwa keempatnya bersifat bias. Teorema 2 dan lemma 3 memberikan  $\hat{\beta}(\lambda, \alpha) = C(\lambda, \alpha)y$ ,  $\hat{g}_\alpha(t) = M(\alpha)y$ ,  $\hat{f}_{\alpha\lambda}(x, t) = N(\lambda, \alpha)y$ , dan  $\hat{\mu}_{\alpha,\lambda}(x, t) = P(\lambda, \alpha)y$ . Terlihat bahwa estimator-estimator  $\hat{\beta}(\lambda, \alpha)$ ,  $\hat{g}_\alpha(t)$ ,  $\hat{f}_{\alpha,\lambda}(x, t)$  dan  $\hat{\mu}_{\alpha,\lambda}(x, t)$  merupakan kelas estimator linear dalam observasi  $y$ .

Estimator campuran regresi Spline dan Kernel  $\hat{\mu}_{\alpha,\lambda}(x, t)$  sangat tergantung pada banyak dan lokasi titik-titik knot  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)'$  dan parameter bandwidth yang optimal. Untuk tujuan ini dapat digunakan berbagai metode. Salah satu metode yang sangat terkenal adalah Generalized Cross Validation (GCV). Fungsi GCV diberikan oleh:

$$G(\lambda, \alpha) = \frac{n^{-1} \|y - \hat{\mu}_{\alpha,\lambda}(x, t)\|^2}{\left(n^{-1} \text{Trace}[\mathbf{I} - \mathbf{N}(\lambda, \alpha) - \mathbf{M}(\alpha)]\right)^2} \quad (4.18)$$

Titik knot optimal  $\lambda_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, \dots, \lambda_{k(opt)})'$  dan parameter bandwidth optimal  $\alpha_{opt}$  diperoleh dari optimasi :  $G(\lambda_{opt}, \alpha_{opt}) = \underset{\lambda, \alpha}{Min} \{G(\lambda, \alpha)\}$ .

#### **4.3. Pemodelan Tingkat Pengangguran Terbuka Propinsi Jawa Barat Menggunakan Campuran Regresi Nonparametrik Spline *Truncated* dan Kernel**

Tingkat Pengangguran Terbuka di propinsi Jawa Barat masih menunjukkan angka yang tinggi untuk Indonesia. TPT yang tinggi menunjukkan bahwa terdapat banyak angkatan kerja yang tidak terserap pada pasar kerja. Sebagaimana ditunjukkan di tabel 4.1, rata-rata TPT untuk semua kabupaten/kota di Jawa Barat adalah sebesar 9,07%. Artinya dari 100 penduduk usia 15 tahun keatas yang tersedia untuk memproduksi barang dan jasa (angkatan kerja) terdapat sebanyak 9 orang yang merupakan penganggura. Selain itu dengan karakteristik yang beragam untuk wilayah kabupaten kota di propinsi Jawa sehingga mengakibatkan beragam pula untuk TPT nya. Pada penelitian ini menggunakan variabel respon (Y) Tingkat Pengangguran Terbuka dengan 7 variabel prediktor yang diduga

mempengaruhi Tingkat Pengangguran Terbuka di Jawa Barat. Analisis pertama yang dipergunakan adalah analisis deskriptif untuk mengetahui karakteristik TPT dengan ketujuh variabel prediktor yang diduga mempengaruhinya. TPT terbesar adalah 14,88%, yaitu Kabupaten Cirebon, sementara untuk TPT terkecil adalah sebesar 5,85%, yaitu Kabupaten Ciamis, dengan tingkat keberagaman untuk setiap kabupaten sebesar 5,118.

Variabel yang diduga mempengaruhi TPT adalah variabel kepadatan penduduk ( $X_1$ ) diperoleh rata-rata 3.710,77, untuk laju pertumbuhan ekonomi ( $X_2$ ) diperoleh rata-rata sebesar 5,51, untuk rata-rata lama sekolah ( $X_3$ ) adalah 8,37 tahun. Variabel yang ke empat adalah upah minimum ( $X_4$ ), dengan rata-rata sebesar Rp 1.335.532, dengan keragaman yang sangat besar yaitu 190.566.569.864. untuk rata-rata persentase penduduk usia kerja di sektor pertanian ( $X_5$ ) diperoleh sebesar 20,76%. Untuk rata-rata persentase panjang jalan dengan kondisi rusak ( $X_6$ ) adalah sebesar 28,17%. Sedangkan variabel terakhir adalah Persentase unit usaha industri kecil, sedang, dan besar per jumlah penduduk usia kerja ( $X_7$ ) dengan rata-rata 1,87.

**Tabel 4.1** Statistika Deskriptif TPT dengan 7 Variabel Prediktor

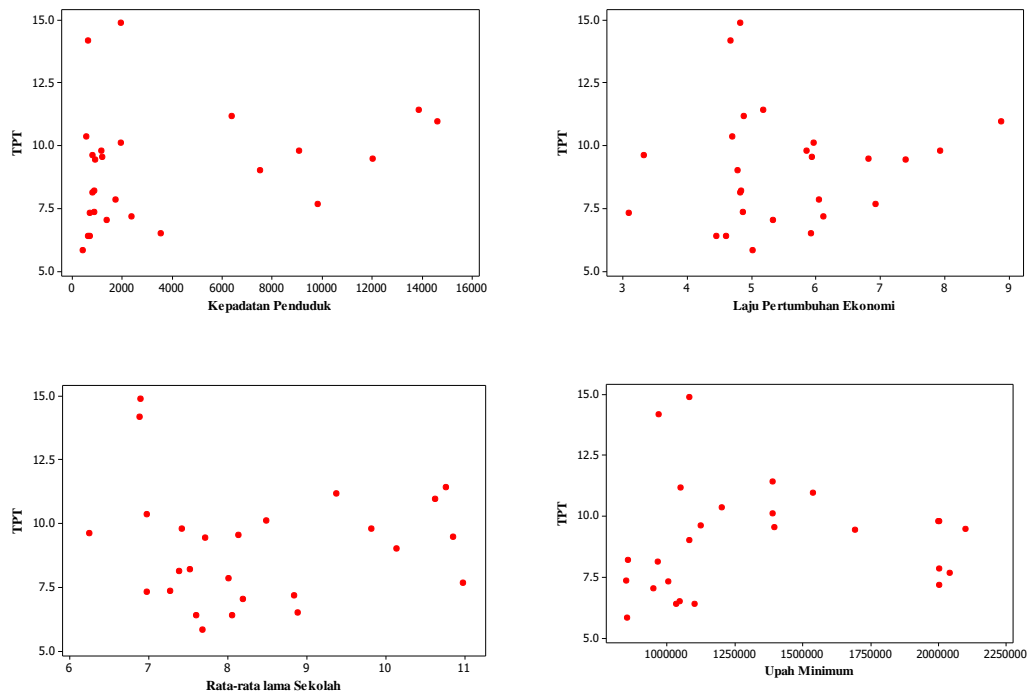
Variabel	Mean	Maksimum	Minimum	Varsians
$Y$	9,07	14,88	5,85	5,118
$X_1$	3.710,77	14.614	422	20.309.253
$X_2$	5,51	8,87	3,10	1,703
$X_3$	8,37	10,98	6,25	1,956
$X_4$	1.335.532	2.100.000	850.000	190.566.569.864
$X_5$	20,76	42,86	0,71	212,598
$X_6$	28,17	62,78	3,14	317,085
$X_7$	1,89	11,71	0,01	7,085

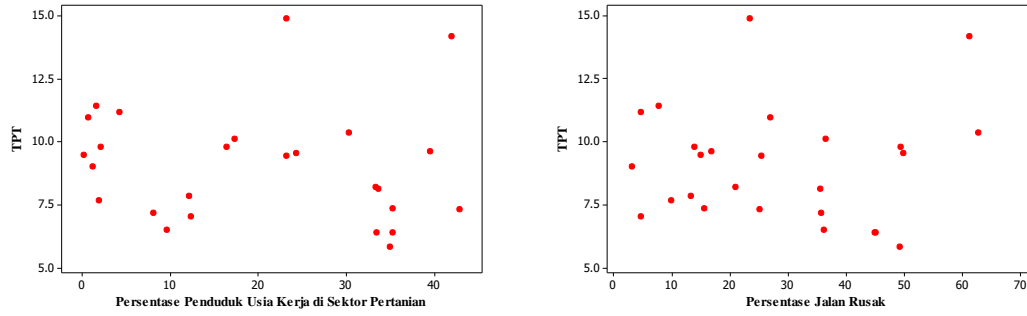
#### 4.3.1 Eksplorasi Data

Sebelum memodelkan TPT secara matematis, perlu diketahui variabel prediktor mana yang cocok dimodelkan dengan spline maupun kernel. Metode yang digunakan terbatas pada *visual checking* dari diagram pencarnya yaitu dengan melihat persebaran titik-titik pada diagram pencar pada Gambar 4.1.

Gambar 4.1 merupakan kumpulan diagram pencar antara variabel respon dan ketujuh prediktornya. Jika dilihat secara menyeluruh, terlihat bahwa pola hubungan antara variabel respon dan prediktor tidak diketahui dan tidak mengikuti pola hubungan tertentu. Namun jika dilihat secara local, dengan pembagian interval tertentu, maka pola hubungan diagram-diagram tersebut dapat diketahui.

Selanjutnya yang akan dilakukan adalah menentukan model hubungan setiap variabel prediktor terhadap variabel respon. Untuk diagram pencar dengan variabel prediktor, rata-rata lama sekolah dan persentase panjang jalan rusak terlihat bahwa model yang tepat adalah spline, dengan komponen polinomial berderajat satu. Karena pada umumnya pola data pada daerah lokalnya membentuk garis lurus namun berubah di interval-interval tertentu. Sedangkan diagram pencar dengan variabel respon kepadatan penduduk, upah minimum, persentase penduduk usia kerja di sektor pertanian, laju pertumbuhan ekonomi, dan Persentase unit usaha industri kecil, sedang, dan besar per jumlah penduduk usia kerja digunakan model kernel.





**Gambar 4.1** Digram Pencar variabel respon TPT dengan ketujuh prediktor

#### 4.3.2 Model Umum Regresi Campuran Nonparametrik Spline dan Kernel

Berdasarkan data dan informasi yang diperoleh dari eksplorasi data, diputuskan pola data yang mengikuti model spline yaitu variabel prediktor rata-rata lama sekolah ( $x_1$ ), persentase panjang jalan rusak ( $x_2$ ), sedangkan pola data yang mengikuti model kernel adalah kepadatan penduduk ( $t_1$ ), laju pertumbuhan ekonomi ( $t_2$ ), upah minimum ( $t_3$ ), persentase penduduk usia kerja di sektor pertanian ( $t_4$ ), dan persentase unit usaha industri kecil, sedang, dan besar per jumlah penduduk usia kerja ( $t_5$ ). Sehingga, model umum regresi campuran nonparametrik spline dan kernel untuk studi kasus TPT adalah sebagai berikut:

$$\begin{aligned}
 y_i &= \sum_{s=1}^2 f_s(x_{si}) + \sum_{k=1}^5 g_k(t_{ki}) + \varepsilon_i, i = 1, 2, 3, \dots, 26 \\
 &= f_1(x_{1i}) + f_2(x_{2i}) + g_1(t_{1i}) + g_2(t_{2i}) + g_3(t_{3i}) + g_4(t_{4i}) + g_5(t_{5i}) + \varepsilon_i \\
 &= \theta_{10} + \theta_{11}x_{1i} + \theta_{12}x_{1i}^2 + \dots + \theta_{1m_1}x_{1i}^{m_1} + \phi_{11}(x_{1i} - \lambda_{11})_+^{m_1} + \dots + \phi_{1r_1}(x_{1i} - \lambda_{1r_1})_+^{m_1} \\
 &\quad + \theta_{20} + \theta_{21}x_{2i} + \theta_{22}x_{2i}^2 + \dots + \theta_{2m_2}x_{2i}^{m_2} + \phi_{21}(x_{2i} - \lambda_{21})_+^{m_2} + \dots + \phi_{2r_2}(x_{2i} - \lambda_{2r_2})_+^{m_2} \\
 &\quad + \sum_{i=1}^{26} \left( \frac{\frac{1}{\alpha_1} K\left(\frac{(t_1 - t_{1i})}{\alpha_1}\right)}{\sum_{j=1}^n \frac{1}{\alpha_1} K\left(\frac{(t_1 - t_{1i})}{\alpha_1}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{\alpha_2} K\left(\frac{(t_2 - t_{2i})}{\alpha_2}\right)}{\sum_{j=1}^n \frac{1}{\alpha_2} K\left(\frac{(t_2 - t_{2i})}{\alpha_2}\right)} \right) y_i
 \end{aligned}$$



$$\begin{aligned}
& + \sum_{i=1}^{26} \left( \frac{\frac{1}{\alpha_3} K\left(\frac{(t_3 - t_{3i})}{\alpha_3}\right)}{\sum_{j=1}^n \frac{1}{\alpha_3} K\left(\frac{(t_3 - t_{3j})}{\alpha_3}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{\alpha_4} K\left(\frac{(t_4 - t_{4i})}{\alpha_4}\right)}{\sum_{j=1}^n \frac{1}{\alpha_4} K\left(\frac{(t_4 - t_{4j})}{\alpha_4}\right)} \right) y_i \\
& + \sum_{i=1}^{26} \left( \frac{\frac{1}{\alpha_5} K\left(\frac{(t_5 - t_{5i})}{\alpha_5}\right)}{\sum_{j=1}^n \frac{1}{\alpha_5} K\left(\frac{(t_5 - t_{5j})}{\alpha_5}\right)} \right) y_i + \varepsilon_i
\end{aligned} \tag{4.19}$$

Dalam bentuk matrik, dapat ditulis  $\mathbf{y}(x, t) = \boldsymbol{\mu}(x, t) + \boldsymbol{\varepsilon}$ . Telah diperoleh di 4.1 model estimator  $\boldsymbol{\mu}(x, t)$ , yaitu:

$$\hat{\boldsymbol{\mu}}(x, t) = \sum_{s=1}^4 \hat{\mathbf{f}}_s(x, t) + \sum_{k=1}^3 \hat{\mathbf{g}}_k(t) = \mathbf{X}(\boldsymbol{\lambda}) \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha) + \mathbf{M}(\alpha) \mathbf{y}, \tag{4.20}$$

dimana nilai dari estimator  $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha) = \left[ (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{X}(\boldsymbol{\lambda})) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y}$ ,

dengan  $\mathbf{X}(\boldsymbol{\lambda}) = [\mathbf{X}_1 \quad \mathbf{X}_2(\boldsymbol{\lambda})]$ .  $\mathbf{X}_1$ ,  $\mathbf{X}_2(\boldsymbol{\lambda})$ , dan  $\mathbf{M}(\alpha)$  berturut-turut diberikan oleh:

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^{m_1} & 1 & x_{21} & \cdots & x_{21}^{m_2} \\ 1 & x_{12} & \cdots & x_{12}^{m_1} & 1 & x_{22} & \cdots & x_{22}^{m_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1.26} & \cdots & x_{1.26}^{m_1} & 1 & x_{2.26} & \cdots & x_{2.26}^{m_2} \end{bmatrix} \tag{4.21}$$

$$\mathbf{X}_2(\boldsymbol{\lambda}) = \begin{bmatrix} (x_{11} - \lambda_{11})_+^{m_1} & \cdots & (x_{11} - \lambda_{1r_1})_+^{m_1} & \cdots & (x_{21} - \lambda_{4r_2})_+^{m_2} \\ (x_{12} - \lambda_{11})_+^{m_1} & \cdots & (x_{12} - \lambda_{1r_1})_+^{m_1} & \cdots & (x_{22} - \lambda_{4r_2})_+^{m_2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (x_{1.26} - \lambda_{11})_+^{m_1} & \cdots & (x_{1.26} - \lambda_{1r_1})_+^{m_1} & \cdots & (x_{2.26} - \lambda_{4r_2})_+^{m_2} \end{bmatrix} \tag{4.22}$$

$$\mathbf{M}(\alpha) = \begin{pmatrix} \sum_{k=1}^5 W_{\alpha_k 1}(t_1) & \sum_{k=1}^5 W_{\alpha_k 2}(t_1) & \cdots & \sum_{k=1}^5 W_{\alpha_k 26}(t_1) \\ \sum_{k=1}^5 W_{\alpha_k 1}(t_2) & \sum_{k=1}^5 W_{\alpha_k 2}(t_2) & \cdots & \sum_{k=1}^5 W_{\alpha_k 26}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{k=1}^5 W_{\alpha_k 1}(t_{26}) & \sum_{k=1}^5 W_{\alpha_k 2}(t_{26}) & \cdots & \sum_{k=1}^5 W_{\alpha_k 26}(t_{26}) \end{pmatrix} \tag{4.23}$$

Karena pada Gambar 4.1 terlihat bahwa pola data pada keempat prediktor spline membentuk garis lurus secara lokal tapi berubah di interval-interval tertentu, maka komponen polinomial untuk semua prediktor tersebut adalah linear ( $m_1 = m_2 = 1$ ). Sehingga, secara berturut-turut persamaan (4.21) dan (4.22) dapat ditulis lebih sederhana menjadi:

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_{11} & 1 & x_{21} \\ 1 & x_{12} & 1 & x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,26} & 1 & x_{2,26} \end{bmatrix} \quad (4.24)$$

$$\mathbf{X}_2(\boldsymbol{\lambda}) = \begin{bmatrix} (x_{11} - \lambda_{11})_+ & \cdots & (x_{11} - \lambda_{1r_1})_+ & \cdots & (x_{21} - \lambda_{2r_2})_+ \\ (x_{12} - \lambda_{11})_+ & \cdots & (x_{12} - \lambda_{1r_1})_+ & \cdots & (x_{22} - \lambda_{2r_2})_+ \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (x_{1,26} - \lambda_{11})_+ & \cdots & (x_{1,26} - \lambda_{1r_1})_+ & \cdots & (x_{2,26} - \lambda_{2r_2})_+ \end{bmatrix} \quad (4.25)$$

#### 4.3.3 Pemilihan Banyaknya Titik-titik Knot dan Lebar *Bandwidth*

Untuk menentukan lokasi dan banyaknya titik knot optimum  $\boldsymbol{\lambda}_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, \dots, \lambda_{k(opt)})'$  dan lebar parameter bandwidth optimal  $\alpha_{opt}$ , digunakan optimasi  $G(\boldsymbol{\lambda}_{opt}, \alpha_{opt}) = \min_{\boldsymbol{\lambda}, \alpha} \{G(\boldsymbol{\lambda}, \alpha)\}$ , dengan  $G$  di persamaan (4.18), dimana  $\mathbf{N}(\boldsymbol{\lambda}, \alpha) = \mathbf{X}(\boldsymbol{\lambda}) \left[ (\mathbf{X}(\boldsymbol{\lambda}))' \mathbf{X}(\boldsymbol{\lambda}) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{I} - \mathbf{M}(\alpha))$ . Langkah awal dalam pemilihan titik knot untuk model spline dilakukan dengan terlebih dahulu mencari nilai GCV optimum secara univariabel untuk masing-masing hubungan respon-prediktor. Kemudian dari nilai tersebut, dibuat selang interval, satu interval ke kiri dan satu ke kanan dengan jarak tertentu dari masing-masing nilai yang diperoleh.

**Tabel 4.2** Perbandingan nilai GCV masing-masing metode secara univariabel

Variabel	Kernel	Spline		
		1 knot	2 knot	3 knot
$X_1$	*5.74	6.50	6.59	6.44
$X_2$	*5.78	7.06	7.51	6.33
$X_3$	5.18	5.72	3.03	*2.87
$X_4$	*2.70	6.52	6.81	6.08
$X_5$	*0.98	6.88	5.74	4.98
$X_6$	5.31	5.85	5.59	*5.10
$X_7$	*1.30	7.03	6.37	5.84

\*Nilai GCV minimum

Asumsi yang digunakan adalah nilai optimum GCV gabungan berada di persekitaran atau bahkan sama dengan nilai GCV univariabel dan multivariabel. Setelah nilai knot dan lebar bandwidth diperoleh, nilai-nilai tersebut dikombinasikan kemudian disubstitusi ke fungsi GCV gabungan. Penentuan banyaknya knot juga berdasarkan GCV. Berdasarkan Tabel 4.2 terlihat bahwa penentuan metode untuk masing-masing variabel telah sesuai dengan penentuan sebelumnya yang menggunakan *scatter plots*. Namun Tabel 4.2 memberikan informasi yang lebih detail yaitu spline menggunakan tiga knot. Karena nilai GCV yang paling minimum dibandingkan dengan nilai GCV yang lain. Selain itu, Nilai GCV multivariabel pada Tabel 4.3 memberikan keputusan yang sama dengan melihat nilai GCV minimumnya. Nilai GCV minimum juga diperoleh ketika variabel rata-rata lama sekolah dan persentase jalan rusak dimodelkan dengan pendekatan spline.

Berdasarkan persamaan 4.23, 4.24, dan 4.25 dan Tabel 4.2, model umum:

$$\begin{aligned}
y_i = & \theta_{10} + \theta_{11}x_{1i} + \phi_{11}(x_{1i} - \lambda_{11})_+ + \phi_{12}(x_{1i} - \lambda_{12})_+ + \phi_{13}(x_{1i} - \lambda_{13})_+ \\
& + \theta_{20} + \theta_{21}x_{2i} + \phi_{21}(x_{2i} - \lambda_{21})_+ + \phi_{22}(x_{2i} - \lambda_{22})_+ + \phi_{23}(x_{2i} - \lambda_{23})_+ \\
& + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_1} K\left(\frac{(t_1 - t_{1i})}{\alpha_1}\right)}{\sum_{j=1}^n \frac{1}{\alpha_1} K\left(\frac{(t_1 - t_{1j})}{\alpha_1}\right)} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_2} K\left(\frac{(t_2 - t_{2i})}{\alpha_2}\right)}{\sum_{j=1}^n \frac{1}{\alpha_2} K\left(\frac{(t_2 - t_{2j})}{\alpha_2}\right)} \right) y_i
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_3} K\left(\frac{(t_3 - t_{3i})}{\alpha_3}\right)}{\sum_{j=1}^n \frac{1}{\alpha_3} K\left(\frac{(t_3 - t_{3j})}{\alpha_3}\right)} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_4} K\left(\frac{(t_4 - t_{4i})}{\alpha_4}\right)}{\sum_{j=1}^n \frac{1}{\alpha_4} K\left(\frac{(t_4 - t_{4j})}{\alpha_4}\right)} \right) y_i \\
& + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_5} K\left(\frac{(t_5 - t_{5i})}{\alpha_5}\right)}{\sum_{j=1}^n \frac{1}{\alpha_5} K\left(\frac{(t_5 - t_{5j})}{\alpha_5}\right)} \right) y_i + \varepsilon_i
\end{aligned}$$

**Tabel 4.3** Perbandingan nilai GCV masing-masing metode secara multivariabel

No	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	GCV
1	Kernel	Kernel	Spline	Kernel	Kernel	Spline	Kernel	3.22
2	Kernel	Spline	Kernel	Spline	Spline	Spline	Kernel	4.07
3	Kernel	Spline	Spline	Spline	Kernel	Kernel	Spline	5.59
4	Spline	Spline	Kernel	Spline	Spline	Kernel	Kernel	6.07
5	Kernel	Spline	Spline	Kernel	Spline	Kernel	Spline	7.64
6	Spline	Kernel	Spline	Spline	Kernel	Spline	Kernel	8.25
7	Kernel	Spline	Spline	Spline	Spline	Kernel	Kernel	8.4
8	Spline	Spline	Kernel	Spline	Kernel	Kernel	Spline	8.46
9	Spline	Spline	Kernel	Kernel	Spline	Kernel	Spline	9.09
10	Spline	Kernel	Kernel	Spline	Spline	Spline	Kernel	13.71
11	Spline	Kernel	Spline	Spline	Spline	Kernel	Kernel	14.09
12	Kernel	Kernel	Spline	Spline	Spline	Spline	Kernel	16.44
13	Spline	Kernel	Spline	Spline	Kernel	Kernel	Spline	22.98
14	Spline	Kernel	Kernel	Kernel	Spline	Spline	Spline	26.85
15	Kernel	Kernel	Spline	Spline	Kernel	Spline	Spline	29.35
16	Kernel	Kernel	Spline	Kernel	Spline	Spline	Spline	30.71
17	Spline	Kernel	Kernel	Spline	Spline	Kernel	Spline	37.11
18	Kernel	Kernel	Kernel	Spline	Spline	Spline	Spline	44.8
19	Kernel	Spline	Kernel	Spline	Spline	Kernel	Spline	47.8
20	Kernel	Kernel	Spline	Spline	Spline	Kernel	Spline	51.08

#### 4.3.4 Model untuk Tiga Titik Knot

Titik knot dan nilai GCV optimum diperlihatkan pada Tabel 4.4. Terlihat bahwa terdapat tiga knot untuk masing-masing variabel.

**Tabel 4.4** Titik Knot dan GCV untuk Model Campuran

No	$\lambda_1$	$\lambda_2$	$\alpha_1$	$\alpha_2$	$\alpha_2$	$\alpha_3$	$\alpha_4$	GCV
1	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0125	3.2206
	6.976	44.481						
	10.792	52.445						
2	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0100	3.2478
	6.976	44.481						
	10.792	52.445						
3	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0075	3.2742
	6.976	44.481						
	10.792	52.445						
4	6.949	23.472	6350.12	4.31	0.0075	0.0075	0.0100	3.2896
	6.976	44.481						
	10.792	52.445						
5	6.955	33.637	6350.12	4.31	0.0075	0.0075	0.0125	3.3100
	6.976	44.481						
	10.792	52.445						
6	6.949	23.472	6350.12	4.31	0.0075	0.0075	0.0075	3.3169
	6.976	44.481						
	10.792	52.445						
7	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0125	3.3193
	6.976	44.481						
	10.792	61.595						
8	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0125	3.3398
	6.976	44.481						
	10.792	60.408						
9	6.955	33.637	6350.12	4.31	0.0075	0.0075	0.0100	3.3418
	6.976	44.481						
	10.792	52.445						
10	6.955	23.472	6350.12	4.31	0.0075	0.0075	0.0100	3.3457
	6.976	44.481						
	10.792	61.595						

Dari Tabel 4.4, terlihat bahwa nilai GCV minimum yang diperoleh adalah 3,2206 dengan titik-titik knot ( $\lambda_{11}$ ,  $\lambda_{12}$ ,  $\lambda_{13}$ ,  $\lambda_{21}$ ,  $\lambda_{22}$ ,  $\lambda_{23}$ ) berturut-turut sama dengan (6.955, 6.976, 10.792, 23.472, 44.481, 52.445) dan bandwidth ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,

$\alpha_4, \alpha_5$ ) yang sama dengan (6350.12, 4.31, 0.0075, 0.0075, 0.0125). Sehingga nilai  $\hat{\beta}(\lambda, \alpha)$  dan  $\hat{\mu}$  berturut-turut adalah:

$$\hat{\beta}(\lambda, \alpha) = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}, \text{ dimana } \hat{\beta}_1 = \begin{bmatrix} 6.365669 \\ -0.75194 \\ 6.365669 \\ 0.000744 \\ 0.064279 \end{bmatrix} \text{ dan } \hat{\beta}_2 = \begin{bmatrix} 1.878450 \\ -28.6414 \\ -0.05368 \\ 0.563072 \\ -3.19290 \end{bmatrix}$$

Dengan demikian estimasi kurva regresi model campuran menjadi:

$$\begin{aligned} \hat{\mu} = & 12.73134 - 0.75194x_{1i} + 0.064279(x_{1i} - 6.955)_+ + 1.878450(x_{1i} - 6.976)_+ \\ & - 28.6414(x_{1i} - 10.792)_+ + 0.000744x_{2i} - 0.05368(x_{2i} - 23.472)_+ \\ & + 0.563072(x_{2i} - 44.481)_+ - 3.19290(x_{2i} - 52.445)_+ \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{6350.12} K\left(\frac{(t_1 - t_{1i})}{6350.12}\right)}{\sum_{j=1}^{26} \frac{1}{6350.12} K\left(\frac{(t_1 - t_{1j})}{6350.12}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{4.31} K\left(\frac{(t_2 - t_{2i})}{4.31}\right)}{\sum_{j=1}^{26} \frac{1}{4.31} K\left(\frac{(t_2 - t_{2j})}{4.31}\right)} \right) y_i \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_3 - t_{3i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_3 - t_{3j})}{0.0075}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_4 - t_{4i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_4 - t_{4j})}{0.0075}\right)} \right) y_i \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0125} K\left(\frac{(t_5 - t_{5i})}{0.0125}\right)}{\sum_{j=1}^{26} \frac{1}{0.0125} K\left(\frac{(t_5 - t_{5j})}{0.0125}\right)} \right) y_i \end{aligned} \quad (4.26)$$

Dari hasil penghitungan menggunakan syntax matlab pada Lampiran 8 diperoleh nilai  $R^2$  sebesar 0,7805. Nilai  $R^2$  ini menunjukkan bahwa variabel yang digunakan dapat menjelaskan model sebesar 78,05%.

#### 4. 4. Interpretasi Model dan Analisis Kebijakan

Model untuk pola data yang mengikuti kernel (kepadatan penduduk, laju pertumbuhan ekonomi, upah minimum, persentase penduduk usia kerja di sektor pertanian, dan persentase unit usaha kecil, sedang, dan besar) tidak dapat diinterpretasi, namun pola data yang mengikuti spline truncated (rata-rata lama sekolah dan persentase panjang jalan rusak) dapat diinterpretasi. Untuk menginterpretasi pola data rata-rata lama sekolah dan persentase panjang jalan rusak, data terlebih dahulu dibagi menjadi 4 bagian untuk setiap variabel tersebut, yaitu data sebelum knot satu, data yang berada di antara knot satu dan knot dua, data yang berada di antara knot dua dan knot tiga, dan data setelah knot tiga. Berikut ini adalah model masing-masing kelompok data tersebut dan interpretasi

##### 1. Model untuk data rata-rata lama sekolah

Dengan mengasumsikan data selain rata-rata lama sekolah konstan, maka diperoleh model umumnya:

$$\begin{aligned}\hat{y}_i &= -0.75194x_{1i} + 0.064279(x_{1i} - 6.955)_+ + 1.878450(x_{1i} - 6.976)_+ \\ &\quad - 28.6414(x_{1i} - 10.792)_+ + c_1 \\ c_1 &= 12.73134 + 0.000744x_{2i} - 0.05368(x_{2i} - 23.472)_+ \\ &\quad + 0.563072(x_{2i} - 44.481)_+ - 3.19290(x_{2i} - 52.445)_+ \\ &\quad + \sum_{i=1}^{26} \left( \frac{\frac{1}{6350.12} K\left(\frac{(t_1 - t_{1i})}{6350.12}\right)}{\sum_{j=1}^{26} \frac{1}{6350.12} K\left(\frac{(t_1 - t_{1j})}{6350.12}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{4.31} K\left(\frac{(t_2 - t_{2i})}{4.31}\right)}{\sum_{j=1}^{26} \frac{1}{4.31} K\left(\frac{(t_2 - t_{2j})}{4.31}\right)} \right) y_i \\ &\quad + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_3 - t_{3i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_3 - t_{3j})}{0.0075}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_4 - t_{4i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_4 - t_{4j})}{0.0075}\right)} \right) y_i \\ &\quad + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0125} K\left(\frac{(t_5 - t_{5i})}{0.0125}\right)}{\sum_{j=1}^{26} \frac{1}{0.0125} K\left(\frac{(t_5 - t_{5j})}{0.0125}\right)} \right) y_i\end{aligned}$$

Untuk interpretasi model ini digunakan fungsi majemuk seperti di bawah ini:

$$\hat{y}_i = \begin{cases} -0.75194x_{1i} + c_1, & x_{1i} \leq 6.955 \\ 0.44706 - 0.6877x_{1i} + c_1, & 6.955 < x_{1i} \leq 6.976 \\ -13.1041 + 1.19075x_{1i} + c_1, & 6.976 < x_{1i} \leq 10.792 \\ 309.085 - 27.45065x_{1i} + c_1, & x_{1i} > 10.792 \end{cases}$$

dengan mengasumsikan data yang lain konstan, maka

- ketika rata-rata lama sekolah naik sebesar 1 tahun dan nilainya kurang dari 6.955 tahun, tingkat pengangguran terbuka akan turun sebesar 0,752 persen.
- ketika rata-rata lama sekolah naik sebesar 1 tahun dan nilainya berada di rentang 6.95-6.98 tahun, tingkat pengangguran terbuka akan turun sebesar 0,688 persen.
- ketika rata-rata lama sekolah naik sebesar 1 tahun dan nilainya berada di rentang 6.976-10.792 tahun, tingkat pengangguran terbuka akan meningkat sebesar 1.9075 persen.
- ketika rata-rata lama sekolah naik sebesar 1 tahun dan nilainya lebih besar 10.792 tahun, tingkat pengangguran terbuka akan menurun sebesar 27.45 persen.

## 2. Model untuk data persentase jalan rusak

Dengan mengasumsikan data selain persentase jalan rusak konstan, maka diperoleh model umumnya:

$$\hat{y}_i = 0.000744x_{2i} - 0.05368(x_{2i} - 23.472)_+ + 0.563072(x_{2i} - 44.481)_+ - 3.19290(x_{2i} - 52.445)_+ + c_2$$

dimana:

$$c_2 = 12.73134 - 0.75194x_{1i} + 0.064279(x_{1i} - 6.955)_+ + 1.878450(x_{1i} - 6.976)_+ - 28.6414(x_{1i} - 10.792)_+ + \sum_{i=1}^{26} \left( \frac{\frac{1}{6350.12} K\left(\frac{(t_1 - t_{1i})}{6350.12}\right)}{\sum_{j=1}^{26} \frac{1}{6350.12} K\left(\frac{(t_1 - t_{1j})}{6350.12}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{4.31} K\left(\frac{(t_2 - t_{2i})}{4.31}\right)}{\sum_{j=1}^{26} \frac{1}{4.31} K\left(\frac{(t_2 - t_{2j})}{4.31}\right)} \right) y_i$$



$$\begin{aligned}
& + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K \left( \frac{(t_3 - t_{3i})}{0.0075} \right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K \left( \frac{(t_3 - t_{3j})}{0.0075} \right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K \left( \frac{(t_4 - t_{4i})}{0.0075} \right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K \left( \frac{(t_4 - t_{4j})}{0.0075} \right)} \right) y_i \\
& + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0125} K \left( \frac{(t_5 - t_{5i})}{0.0125} \right)}{\sum_{j=1}^{26} \frac{1}{0.0125} K \left( \frac{(t_5 - t_{5j})}{0.0125} \right)} \right) y_i
\end{aligned}$$

Untuk interpretasi model ini digunakan fungsi majemuk seperti di bawah ini:

$$\hat{y}_i = \begin{cases} 0.000744x_{2i} + c_2, & x_{2i} \leq 23.472 \\ 1.25998 - 0.05294x_{2i} + c_2, & 23.472 < x_{2i} \leq 44.481 \\ -23.786 + 0.5101x_{2i} + c_2, & 44.481 < x_{2i} \leq 52.445 \\ 143.6656 - 2.6828x_{2i} + c_2, & x_{2i} > 52.445 \end{cases}$$

dengan mengasumsikan data yang lain konstan, maka:

- ketika Persentase jalan rusak naik sebesar satu persen dan nilainya kurang dari 23.47 persen, tingkat pengangguran terbuka akan naik sebesar 0,000744 persen.
- ketika Persentase jalan rusak naik sebesar satu persen dan nilainya berada di rentang 23.47-44.48 persen, tingkat pengangguran terbuka akan turun sebesar 0,05294 persen.
- ketika Persentase jalan rusak naik sebesar satu persen dan nilainya berada di rentang 44.48-52.445 persen, tingkat pengangguran terbuka akan meningkat sebesar 0.5101 persen.
- ketika Persentase jalan rusak naik sebesar satu persen dan nilainya lebih besar 52.445 persen, tingkat pengangguran terbuka akan menurun sebesar 2.6828 persen.

Dari Tabel 4.5 terlihat bahwa semakin tinggi rata-rata lama sekolah suatu kabupaten, akan memberikan nilai TPT yang semakin kecil. Ini telah sesuai dengan teori TPT yang menyatakan pengangguran berbanding terbalik dengan pendidikan. Keenam variabel lainnya juga telah memenuhi teori sebagaimana

International ILO mendefinisikan beberapa peubah yang berpengaruh terhadap TPT, yaitu kependudukan, pendidikan, upah tenaga kerja, PDRB, banyaknya pekerjaan di sektor formal dan informal, infrastruktur, serta sarana dan prasarana yang tersedia di suatu wilayah. Indikator-indikator tersebut menjadi acuan pemerintah dalam mengeluarkan kebijakan pembukaan lapangan kerja baru.

Penjelasan lebih lanjut Tabel 4.5, adalah sebagai berikut:

- a. Optimis. Nilai TPT diharapkan kecil yang diperoleh dari mensubsitusi nilai-nilai setiap indikator ke model estimasi yang diperoleh. Nilai TPT ini merupakan nilai yang sangat diharapkan tercapai melalui nilai setiap variabel yang ditawarkan.
- b. Tengah. Nilai TPT diharapkan sedang-sedang saja yang diperoleh dari mensubsitusi nilai-nilai setiap indikator ke model estimasi yang diperoleh. Nilai TPT ini merupakan nilai yang cukup diharapkan tercapai melalui nilai setiap variabel yang ditawarkan.
- c. Pesimis. Nilai TPT tinggi yang diperoleh dari mensubsitusi nilai-nilai setiap indikator ke model estimasi yang diperoleh. Nilai TPT ini merupakan nilai yang tidak diharapkan tercapai melalui nilai setiap variabel yang ditawarkan.

**Tabel 4.5** Analisis Kebijakan TPT

	No	$x_1$	$x_2$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$\hat{y}$
Optimis	1	11.40	4.00	422	8.87	2100000	42.86	11.71	0.07
	2	11.35	4.60	2196	8.15	1943750	37.53	10.25	0.95
	3	11.30	5.40	3970	7.43	1787500	32.20	8.78	1.85
Tengah	1	11.00	19.60	5744	6.71	1631250	26.86	7.32	7.72
	2	10.9	22.90	7518	5.99	1475000	21.53	5.86	9.60
	3	10.8	24.20	9292	5.26	1318750	16.20	4.40	11.24
Pesimis	1	6.9	45.40	11066	4.54	1162500	10.87	2.93	12.49
	2	6.8	48.20	12840	3.82	1006250	5.53	1.47	12.86
	3	6.7	51.80	14614	3.10	850000	0.20	0.01	13.73

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1. Data dan Format Data yang Digunakan

Kabupaten/Kota	y	x1	x2	x3	x4	x5	x6	x7
Kab. Bogor	7.87	1736	6.04	8.01	2002000	12.15	13.22	0.65
Kab. Sukabumi	10.38	579	4.7	6.97	1201020	30.29	62.78	1.47
Kab. Cianjur	14.18	619	4.67	6.88	970000	41.96	61.19	0.12
Kab. Bandung	10.12	1939	5.96	8.49	1388333	17.3	36.48	0.86
Kab. Garut	8.14	809	4.82	7.39	965000	33.63	35.51	0.98
Kab. Tasikmalaya	6.4	636	4.46	7.6	1035000	35.23	44.84	0.18
Kab. Ciamis	5.85	422	5.02	7.68	854075	34.98	49.24	0.18
Kab. Kuningan	8.22	877	4.84	7.52	857000	33.32	20.91	0.50
Kab. Cirebon	14.88	1954	4.83	6.9	1081300	23.19	23.39	1.19
Kab. Majalengka	7.35	871	4.87	7.27	850000	35.23	15.51	1.25
Kab. Sumedang	6.41	721	4.6	8.06	1100000	33.46	45.10	0.99
Kab. Indramayu	9.63	800	3.33	6.25	1125000	39.57	16.78	0.30
Kab. Subang	7.34	691	3.1	6.98	1005000	42.86	25.13	0.48
Kab. Purwakarta	9.45	907	7.39	7.71	1693167	23.25	25.42	2.73
Kab. Karawang	9.8	1163	7.92	7.42	2000000	16.41	49.36	0.95
Kab. Bekasi	7.17	2365	6.11	8.84	2002000	8.09	35.63	0.78
Kab. Bandung Bar	9.54	1190	5.94	8.14	1396399	24.34	49.83	0.01
Kota Bogor	9.8	9067	5.86	9.82	2002000	2.06	13.85	1.84
Kota Sukabumi	11.18	6368	4.88	9.37	1050000	4.22	4.71	7.09
Kota Bandung	10.97	14614	8.87	10.63	1538703	0.71	26.99	0.91
Kota Cirebon	9.02	7513	4.79	10.14	1082500	1.23	3.14	6.60
Kota Bekasi	9.5	12035	6.81	10.85	2100000	0.2	15.00	0.84
Kota Depok	7.69	9838	6.92	10.98	2042000	1.92	9.80	1.14
Kota Cimahi	11.43	13859	5.18	10.76	1388333	1.62	7.68	2.34
Kota Tasikmalaya	6.52	3534	5.92	8.89	1045000	9.57	36.12	3.15
Kota Banjar	7.04	1373	5.34	8.19	950000	12.36	4.72	11.71
Keterangan: y: Tingkat Pengangguran Terbuka x1: Kepadatan Penduduk x2: Laju Pertumbuhan ekonomi x3: Rata-rata Lama Sekolah x4: Upah minimum x5: Persentase penduduk usia kerja di sektor pertanian x6: persentase panjang jalan rusak x7: Persentase unit usaha industri kecil, sedang, dan besar per jumlah penduduk usia kerja								

## Lampiran 2. Struktur Data pada Matlab

```
datarismal=[7.87 1736 6.04 8.01 2002000 12.15 13.22 0.65;  
10.38 579 4.7 6.97 1201020 30.29 62.78 1.47;  
14.18 619 4.67 6.88 970000 41.96 61.19 0.12;  
10.12 1939 5.96 8.49 1388333 17.3 36.48 0.86;  
8.14 809 4.82 7.39 965000 33.63 35.51 0.98;  
6.4 636 4.46 7.6 1035000 35.23 44.84 0.18;  
5.85 422 5.02 7.68 854075 34.98 49.24 0.18;  
8.22 877 4.84 7.52 857000 33.32 20.91 0.50;  
14.88 1954 4.83 6.9 1081300 23.19 23.39 1.19;  
7.35 871 4.87 7.27 850000 35.23 15.51 1.25;  
6.41 721 4.6 8.06 1100000 33.46 45.10 0.99;  
9.63 800 3.33 6.25 1125000 39.57 16.78 0.30;  
7.34 691 3.1 6.98 1005000 42.86 25.13 0.48;  
9.45 907 7.39 7.71 1693167 23.25 25.42 2.73;  
9.8 1163 7.92 7.42 2000000 16.41 49.36 0.95;  
7.17 2365 6.11 8.84 2002000 8.09 35.63 0.78;  
9.54 1190 5.94 8.14 1396399 24.34 49.83 0.01;  
9.8 9067 5.86 9.82 2002000 2.06 13.85 1.84;  
11.18 6368 4.88 9.37 1050000 4.22 4.71 7.09;  
10.97 14614 8.87 10.63 1538703 0.71 26.99 0.91;  
9.02 7513 4.79 10.14 1082500 1.23 3.14 6.60;  
9.5 12035 6.81 10.85 2100000 0.2 15.00 0.84;  
7.69 9838 6.92 10.98 2042000 1.92 9.80 1.14;  
11.43 13859 5.18 10.76 1388333 1.62 7.68 2.34;  
6.52 3534 5.92 8.89 1045000 9.57 36.12 3.15;  
7.04 1373 5.34 8.19 950000 12.36 4.72 11.71];
```

### Lampiran 3. Fungsi Penentuan Nilai GCV Univariabel (Kernel)

```
function [phimin]=syntaxkernel(intkernel)
data=datarismal;
y=data(:,1); %variabel y
xk=[data(:,2) data(:,3) data(:,4) data(:,5) data(:,6) data(:,7)
data(:,8)]; %variabel kernel
nk=length(xk(1,:)); %banyaknya variabel spline
n=length(xk(:,1)); %jumlah observasi

%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas

for z=1:nk
x=xk(:,z);
selisih=max(x)-min(x);
interval=selisih/(intkernel+1);
phi=0.01:interval:selisih;
s=1;
nn=length(phi);
for e=1:nn
vdiag=diag(x);
V=matnn*vdiag;
Z=(V'-V)/phi(e);
K=1/sqrt(2*pi)*exp(-1/2*Z.^2); % fungsi kernel gaussian
KZ=(1/phi(e))*K;
Wpenyebut=diag(1/n.*KZ*ones(n,1))*matnn;
Vphi=1/n*KZ./Wpenyebut ; % penimbang V(phi).1
% estimasi parameter
teta=Vphi;
yhat=teta*y;

% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSE=SSE/(n-2);
R2=SSR/SST;

% menghitung GCV
GCV(e)=(MSE)/(1/n*sum(diag(matiden-Vphi).^2));
end
value=0;
value=[phi' GCV'];
value(1:10,:);
sortGCV=sortrows(value,2);
phimin(z,:)=sortGCV(1,:);
end
phi=phimin(:,1);
end
```

#### Lampiran 4. Fungsi Penentuan Nilai GCV Univariabel (Spline Satu Knot)

```
function [knotmin]=syntaxspline(intspline)
data=datarismal;
y=data(:,1); %variabel y
xs=[data(:,2) data(:,3) data(:,4) data(:,5) data(:,6) data(:,7)
data(:,8)]; %variabel spline
ns=length(xs(1,:)); %banyaknya variabel spline
n=length(xs(:,1)) ; %jumlah observasi

%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas

for z=1:ns
x=xs(:,z);
selisih=max(x)-min(x);
interval=selisih/(intspline+1);
komspline(1,:)=min(x):interval:max(x);
s=1;
for a=1:intspline
k(1)=komspline(1,a);
r(s,:)=k;
s=s+1;
end
nn=length(r(:,1));
for e=1:nn
k=r(e,:);
X=0;
X=[matn x x-k(1)];

vv=length(X(1,:));
for i=1:vv
for j=1:n
if X(j,i)<0
X(j,i)=0;
else
X(j,i)=X(j,i);
end
end
end

% estimasi parameter
teta=pinv(X'*X)*X'*y;
A=X*pinv(X'*X)*X';
yhat=X*teta;

% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSE=SSE/(n-vv-2);
R2=SSR/SST;
% menghitung GCV
GCV(e)=(MSE)/(1/n*sum(diag(matiden-A).^2));
```

```
end
value=0;
value=[r GCV'];
value(1:10,:);
sortGCV=sortrows(value,2);
knotmin(z,:)=sortGCV(1,:);
end
```



## Lampiran 5. Fungsi Penentuan Nilai GCV Univariabel (Spline Dua Knot)

```
function [knotmin]=syntaxspline2(intspline)
data=datarismal;
y=data(:,1); %variabel y
xs=[data(:,2) data(:,3) data(:,4) data(:,5) data(:,6) data(:,7)
data(:,8)]; %variabel spline
ns=length(xs(1,:)); %banyaknya variabel spline
n=length(xs(:,1)) ; %jumlah observasi

%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas

for z=1:ns
x=xs(:,z);
selisih=max(x)-min(x);
interval=selisih/(intspline+1);
komspline(1,:)=min(x):interval:max(x);
s=1;
for a=1:intspline
k(1)=komspline(1,a);
for b=(a+1):(intspline+1)
k(2)=komspline(1,b);
r(s,:)=k;
s=s+1;
end
end
nn=length(r(:,1));
for e=1:nn
k=r(e,:);
X=0;
X=[matn x x-k(1) x-k(2)];

vv=length(X(1,:));
for i=1:vv
for j=1:n
if X(j,i)<0
X(j,i)=0;
else
X(j,i)=X(j,i);
end
end
end

% estimasi parameter
teta=pinv(X'*X)*X'*y;
A=X*pinv(X'*X)*X';
yhat=X*teta;

% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSE=SSE/(n-vv-2);
R2=SSR/SST;
```

```
% menghitung GCV
GCV(e)=(MSE)/(1/n*sum(diag(matiden-A).^2));
end
value=0;
value=[r GCV'];
value(1:10,:);
sortGCV=sortrows(value,3);
knotmin(z,:)=sortGCV(1,:);
end
```

## Lampiran 6. Fungsi Penentuan Nilai GCV Univariabel (Spline Tiga Knot)

```
function [knotmin]=syntaxspline3(intspline)
data=datarismal;
y=data(:,1); %variabel y
xs=[data(:,2) data(:,3) data(:,4) data(:,5) data(:,6) data(:,7)
data(:,8)]; %variabel spline
ns=length(xs(1,:)); %banyaknya variabel spline
n=length(xs(:,1)) ; %jumlah observasi

%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas

for z=1:ns
x=xs(:,z);
selisih=max(x)-min(x);
interval=selisih/(intspline+1);
komspline(1,:)=min(x):interval:max(x);
s=1;
for a=1:intspline
k(1)=komspline(1,a);
for b=(a+1):(intspline+1)
k(2)=komspline(1,b);
for c=(b+1):(intspline+2)
k(3)=komspline(1,c);
r(s,:)=k;
s=s+1;
end
end
end
nn=length(r(:,1));
for e=1:nn
k=r(e,:);
X=0;
X=[matn x x-k(1) x-k(2) x-k(3)];

vv=length(X(1,:));
for i=1:vv
for j=1:n
if X(j,i)<0
X(j,i)=0;
else
X(j,i)=X(j,i);
end
end
end

% estimasi parameter
teta=pinv(X'*X)*X'*y;
A=X*pinv(X'*X)*X';
yhat=X*teta;
```

```

% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSE=SSE/(n-vv-2);
R2=SSR/SST;

% menghitung GCV
GCV(e)=(MSE)/(1/n*sum(diag(matiden-A).^2));
end
value=0;
value=[r GCV'];
value(1:10,:);
sortGCV=sortrows(value,4);
knotmin(z,:)=sortGCV(1,:);
end

```

## Lampiran 7. Fungsi Penentuan Nilai GCV Multivariabel Empat Spline Tiga Knot dan Tiga Kernel

```

clc;
clear;
data=datarismal;
y=data(:,1); %variabel y
xs=[data(:,2) data(:,4) data(:,5) data(:,6)]; %variabel spline
xk=[data(:,3) data(:,7) data(:,8)]; %variabel kernel
ns=length(xs(1,:)); %banyaknya variabel spline
nk=length(xk(1,:)) ; %banyaknya variabel spline
n=length(xk) ; %jumlah observasi
intspline=2;
intkernel=2;

%seq spline
for i=1:ns
    selisih=(max(xs(:,i))-min(xs(:,i)));
    interval=selisih/(intspline+1);
    komspline(i,:)=min(xs(:,i)):interval:max(xs(:,i));
end
%seq kernel
for i=1:nk
    selisih=(max(xk(:,i))-min(xk(:,i)));
    interval=selisih/(intkernel+1);
    komkernel(i,:)=0.01:interval:selisih;
end

%kombinasi knot & bandwitch
s=1;
interval=intspline-1;
for a=1:intspline
    k(1)=komspline(1,a);
    for b=(a+1):(intspline+1)
        k(2)=komspline(1,b);
        for c=(b+1):(intspline+2)
            k(3)=komspline(1,c);
            for d=1:intspline
                k(4)=komspline(2,d);
                for e=(d+1):(intspline+1)
                    k(5)=komspline(2,e);
                    for f=(e+1):(intspline+2)
                        k(6)=komspline(2,f);
                        for g=1:intspline
                            k(7)=komspline(3,g);
                            for h=(g+1):(intspline+1)
                                k(8)=komspline(3,h);
                                for i=(h+1):(intspline+2)
                                    k(9)=komspline(3,i);
                                    for j=1:intspline
                                        k(10)=komspline(4,j);
                                        for l=(j+1):(intspline+1)
                                            k(11)=komspline(4,l);
                                            for m=(l+1):(intspline+2)
                                                k(12)=komspline(4,m);
                                                for o=1:(intkernel+1)

```

```
k(13)=komkernel(1,o);
for p=1:(intkernel+1)
    k(14)=komkernel(2,p);
    for q=1:(intkernel+1)
        k(15)=komkernel(3,q);
        r(s,:)=k;
        s=s+1
    end
end
end
end
end
end
end
end
end
end
end
end
end
end
%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas
nn=length(r(:,1));
for v=1:nn

    %membuat matrix knot
    kn=r(1,1:12);
    c=1;
    k=0;
    for b=1:ns
        d=c+2;
        k(1:3,b)=kn(c:d)';
        c=d+1;
    end
    phi=r(1,13:15)';

% desain matriks Z(lambda) pada spline
X=0;
x0=[matn xs];
X=x0;
for i=1:ns
    a=0;
    a=[xs(:,i)-k(1,i) xs(:,i)-k(2,i) xs(:,i)-k(3,i)];
    X=[X a];
end
vv=ns*3+(1+ns);
for i=1:vv
    for j=1:n
        if X(j,i)<0
            X(j,i)=0;
        else
            X(j,i)=X(j,i);
        end
    end
end
```

```

end
end

sumvphi=zeros(n,n);
for i=1:nk
    vdiag=diag(xk(:,i));
    V=matnn*vdiag;
    Z=(V'-V)/phi(i);
    K=1/sqrt(2*pi)*exp(-1/2*Z.^2);           % fungsi kernel gaussian
    KZ=(1/phi(i))*K;
    Wpenyebut=diag(1/n.*KZ*ones(n,1))*matnn;
    Vphi=1/n*KZ./Wpenyebut ;                 % penimbang V(phi).1
    sumvphi=sumvphi+Vphi;                     %nilai kernel untuk setiap
variabel
end;
D=sumvphi/nk;

% estimasi parameter
C=pinv(X'*X)*X'*(matiden-D);
beta=C*y;
A=X*C;
B=A+D;
yhat=B*y;

% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSR=SSR/(ns+nk);
MSE=SSE/(n-(ns+nk)-1);
R2=SSR/SST;
e=y-yhat;

% menghitung GCV
GCV(v)=(MSE)/(1/n*sum(diag(matiden-A-D).^2));
end

value=[r GCV'];
sortGCV=sortrows(value,16);
GCVmin=sortGCV(1,16)

```

## Lampiran 8. Fungsi Penentuan Nilai GCV Multivariabel dua Spline Tiga Knot dan lima Kernel dengan Nilai Knot dan Bandwidth Diberikan

```
function [terbaik,teta]=syntax(knotphi)
nkomb=length(knotphi(:,1));
data=datarismal;
y=data(:,1); %variabel y
xs=[data(:,4) data(:,7)]; %variabel spline
xk=[data(:,2) data(:,3) data(:,5) data(:,6) data(:,8)];
%variabel kernel
nk=length(xk(1,:)); %banyaknya variabel spline
ns=length(xs(1,:)); %banyaknya variabel spline
n=length(xs(:,1)); %jumlah observasi
hh=1;
%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas
for aa=1:nkomb
    kb(1)=knotphi(aa,1);
    for bb=1:nkomb
        kb(2)=knotphi(bb,2);
        for c1=1:nkomb
            kb(3)=knotphi(c1,3);
            for c2=1:nkomb
                kb(4)=knotphi(c2,4);
                for c3=1:nkomb
                    kb(5)=knotphi(c3,5);
                    for dd=1:nkomb
                        kb(6)=knotphi(dd,6);
                        for ee=1:nkomb
                            kb(7)=knotphi(ee,7);
                            for f1=1:nkomb
                                kb(8)=knotphi(f1,8);
                                for f2=1:nkomb
                                    kb(9)=knotphi(f2,9);
                                    for f3=1:nkomb
                                        kb(10)=knotphi(f3,10);
                                        for gg=1:nkomb

kb(11)=knotphi(gg,11);
knot1=[kb(3:5);kb(8:10)];
%spline
k=knot1(1:2,1:3);
X=0;
x0=[matn xs(:,1) matn xs(:,2)];
X=x0;
for i=1:ns
    a=0;
    a=[xs(:,i)-k(i,1) xs(:,i)-k(i,2) xs(:,i)-k(i,3)];
    X=[X a];
end

vv=length(X(1,:));
for i=1:vv
    for j=1:n
```



```

        if X(j,i)<0
            X(j,i)=0;
        else
            X(j,i)=X(j,i);
        end
    end
end

phi=[kb(1) kb(2) kb(6) kb(7) kb(11)];
%kernel
for e=1:(length(phi))
    vdiag=diag(xk(:,e));
    V=matnn*vdiag;
    Z=(V'-V)/phi(e);
    K=1/sqrt(2*pi)*exp(-1/2*Z.^2); % fungsi kernel gaussian
    KZ=(1/phi(e))*K;
    Wpenyebut=diag(1/n.*KZ*ones(n,1))*matnn;
    Vphi=1/n*KZ./Wpenyebut ; % penimbang V(phi).1
end
D=Vphi/nk;
% estimasi parameter
C=pinv(X'*X)*X'*(matiden-D);
teta=C*y;
A=X*C;
B=A+D;
yhat=B*y;
% evaluasi Model
SST=sum((y-mean(y)).^2);
SSR=sum((yhat-mean(y)).^2);
SSE=sum((y-yhat).^2);
MSR=SSR/(ns+nk);
MSE=SSE/(n-(ns+nk)-1);
R2=SSR/SST;
e=y-yhat;

% menghitung GCV
GCVoptimum=(MSE)/(1/n*sum(diag(matiden-A-D).^2));
hasil(hh,:)= [kb GCVoptimum];
hh=hh+1;


                                end
                            end
                        end
                    end
                end
            end
        end
    end
end
xlswrite('belum urut.xlsx', hasil);
sortGCV=sortrows(hasil,12);
xlswrite('GCVsort.xlsx', sortGCV);
terbaik=sortGCV(1:10,1:12);
end
```

## Lampiran 9. Syntax Untuk Analisis Kebijakan

```
function [y]=syntaxkebijakan(knotphi,teta)
datakebijakan=[422 8.87 5 2100000 42.86 4.0 11.70825579
2196 8.14875 11.4 1943750 37.5275 4.6 10.24570401
3970 7.4275 11.3 1787500 32.195 5.4 8.783152224
5744 6.70625 11 1631250 26.8625 19.6 7.320600441
7518 5.985 6.1 1475000 21.53 22.9 5.858048659
9292 5.26375 8.8 1318750 16.1975 24.2 4.395496876
11066 4.5425 9.2 1162500 10.865 60.4 2.932945094
12840 3.82125 10.6 1006250 5.5325 57.2 1.470393311
14614 3.1 6.9 850000 0.2 54.8 0.007841529];
nkomb=length(knotphi(:,1));
data=datakebijakan;
xs=[data(:,3) data(:,6)]; %variabel spline
xk=[data(:,1) data(:,2) data(:,4) data(:,5) data(:,7)];
%variabel kernel
nk=length(xk(1,:)); %banyaknya variabel spline
ns=length(xs(1,:)); %banyaknya variabel spline
n=length(xs(:,1)); %jumlah observasi
hh=1;
%matrix
matnn=ones(n,n); %membuat matrix 1 nxn
matn=ones(n,1); %membuat vector 1 nx1
matiden=eye(n); %membuat matrix identitas
for aa=1:nkomb
    kb(1)=knotphi(aa,1);
    for bb=1:nkomb
        kb(2)=knotphi(bb,2);
        for c1=1:nkomb
            kb(3)=knotphi(c1,3);
            for c2=1:nkomb
                kb(4)=knotphi(c2,4);
                for c3=1:nkomb
                    kb(5)=knotphi(c3,5);
                    for dd=1:nkomb
                        kb(6)=knotphi(dd,6);
                        for ee=1:nkomb
                            kb(7)=knotphi(ee,7);
                            for f1=1:nkomb
                                kb(8)=knotphi(f1,8);
                                for f2=1:nkomb
                                    kb(9)=knotphi(f2,9);
                                    for f3=1:nkomb
                                        kb(10)=knotphi(f3,10);
                                        for gg=1:nkomb
                                            kb(11)=knotphi(gg,11);
                                        end
                                    end
                                end
                            end
                        end
                    end
                end
            end
        end
    end
    knot1=[kb(3:5);kb(8:10)];
    %spline
    k=knot1(1:2,1:3);
    X=0;
    x0=[matn xs(:,1) matn xs(:,2)];
    X=x0;
    for i=1:ns
        a=0;
        a=[xs(:,i)-k(i,1) xs(:,i)-k(i,2) xs(:,i)-k(i,3)];
```



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan pembahasan pada Bab 4, maka kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

1. Diberikan model campuran regresi nonparametrik sebagai berikut :

$$y_i = \mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}) + \varepsilon_i, \text{ untuk } i = 1, 2, \dots, n$$

Jika model regresi tersebut dinotasikan dalam bentuk matriks maka diperoleh bentuk :

$$\mathbf{y} = \boldsymbol{\mu}(x, t) + \boldsymbol{\varepsilon} = \sum_{s=1}^p \mathbf{f}_s(x) + \sum_{k=1}^q \mathbf{g}_k(t) + \boldsymbol{\varepsilon}.$$

Komponen  $\mathbf{f}_s(x)$  didekati oleh regresi spline aditif dengan prediktor sebanyak  $p$  sementara  $\mathbf{g}_k(t)$  didekati dengan regresi kernel aditif dengan prediktor sebanyak  $q$ . Error  $\boldsymbol{\varepsilon}$  diasumsikan berdistribusi normal dengan mean nol dan varians konstan. Estimator diperoleh menggunakan *Ordinary Least Square* (OLS):

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}, \alpha) = \left[ (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{X}(\boldsymbol{\lambda})) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' [\mathbf{I} - \mathbf{M}(\alpha)] \mathbf{y} = \mathbf{C}(\boldsymbol{\lambda}, \alpha) \mathbf{y}$$

Untuk model dugaan campuran regresi nonparametrik spline *truncated* dan kernel adalah sebagai berikut:

$$\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t) = \hat{\mathbf{f}}_{\alpha \lambda}(x, t) + \hat{\mathbf{g}}_{\alpha}(t) = [\mathbf{N}(\boldsymbol{\lambda}, \alpha) + \mathbf{M}(\alpha)] \mathbf{y}$$

dimana:

$$\mathbf{N}(\boldsymbol{\lambda}, \alpha) = \mathbf{X}(\boldsymbol{\lambda}) \left[ (\mathbf{X}(\boldsymbol{\lambda}))' \mathbf{X}(\boldsymbol{\lambda}) \right]^{-1} (\mathbf{X}(\boldsymbol{\lambda}))' (\mathbf{I} - \mathbf{M}(\alpha))$$

2. Estimator  $\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t)$  memiliki sifat bias. Karena  $E(\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t)) \neq \boldsymbol{\mu}_{\alpha, \lambda}(x, t)$ . Namun diperoleh  $\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t) = \mathbf{P}(\boldsymbol{\lambda}, \alpha) \mathbf{y}$ . Terlihat bahwa estimator  $\hat{\boldsymbol{\mu}}_{\alpha, \lambda}(x, t)$  merupakan kelas estimator linear dalam observasi  $\mathbf{y}$ .

3. Pemodelan Tingkat Pengangguran Terbuka Provinsi Jawa Barat tahun 2013 menggunakan estimator campuran kernel dan spline *truncated* dalam regresi noparametrik multivariabel adalah sebagai berikut:

$$\begin{aligned} \hat{\mu} = & 12.73134 - 0.75194x_{1i} + 0.064279(x_{1i} - 6.955)_+ + 1.878450(x_{1i} - 6.976)_+ \\ & - 28.6414(x_{1i} - 10.792)_+ + 0.000744x_{2i} - 0.05368(x_{2i} - 23.472)_+ \\ & + 0.563072(x_{2i} - 44.481)_+ - 3.19290(x_{2i} - 52.445)_+ \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{6350.12} K\left(\frac{(t_1 - t_{1i})}{6350.12}\right)}{\sum_{j=1}^{26} \frac{1}{6350.12} K\left(\frac{(t_1 - t_{1j})}{6350.12}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{4.31} K\left(\frac{(t_2 - t_{2i})}{4.31}\right)}{\sum_{j=1}^{26} \frac{1}{4.31} K\left(\frac{(t_2 - t_{2j})}{4.31}\right)} \right) y_i \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_3 - t_{3i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_3 - t_{3j})}{0.0075}\right)} \right) y_i + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0075} K\left(\frac{(t_4 - t_{4i})}{0.0075}\right)}{\sum_{j=1}^{26} \frac{1}{0.0075} K\left(\frac{(t_4 - t_{4j})}{0.0075}\right)} \right) y_i \\ & + \sum_{i=1}^{26} \left( \frac{\frac{1}{0.0125} K\left(\frac{(t_5 - t_{5i})}{0.0125}\right)}{\sum_{j=1}^{26} \frac{1}{0.0125} K\left(\frac{(t_5 - t_{5j})}{0.0125}\right)} \right) y_i \end{aligned}$$

Nilai  $R^2$  model tersebut sebesar 0,7805. Ini artinya variabel yang digunakan dapat menjelaskan model sebesar 78,05%.

## 5.2 Saran

Berikut ini adalah saran yang dapat disampaikan berdasarkan hasil analisis dan pembahasan yang dilakukan pada penelitian ini untuk pihak instansi terkait dan untuk kelanjutan penelitian di masa akan datang.

1. Dalam memutuskan suatu pola hubungan dapat dimodelkan spline ataupun kernel, hendaknya digunakan metode lain. Baik secara matematis maupun nonmatematis. Kasus ini bisa digunakan sebagai studi lanjut

2. Sebagaimana dijelaskan di pendahuluan, Terdapat banyak pendekatan estimasi nonparametrik, penulis menyarankan untuk mempertimbangkan estimator yang lain untuk digabungkan sebagian bahkan seluruhnya untuk menentukan pemodelan awal.
3. Karena keterbatasan pemrograman, penulis hanya menggunakan maksimal tiga titik knot. Untuk penelitian selanjutnya, sebaiknya dicoba lebih banyak titik knot. Derajat polinomial juga hanya terbatas pada derajat satu, yaitu linear. Sebaiknya dipertimbangkan derajat polinomial yang lebih.

*(Halaman ini sengaja dikosongkan)*

## Daftar Pustaka

- Alghofari. (2010). *Analisis Tingkat Pengangguran di Indonesia Tahun 1980-2007*. Universitas Diponegoro: Semarang
- Amato, U., Antoniadis, A., Feis, D. I. (2002). Fourier Series Approximation of Separable Models. *Journal of Computational and Applied Mathematics*, 146: 459-479.
- Amato, U., De-Canditiis D. (2001). Convergence in Probability of the Mallows and GCV Wavelet and Fourier Regularization Methods. *Statistics & Probability Letters*, 54: 325-329.
- Angelini, C., De-Canditiis, D. dan Leblanc, F. (2003). Wavelet Regression Estimation in Nonparametric Mixed Effect Models. *Journal of Multivariate Analysis*, 85: 267-291.
- Antoniadis, A., Bigot, J. dan Sapatinas, T. (2001). Wavelet Estimators in Nonparametric Regression, a Comparative Simulation Study. *Journal of Statistical Software*, 61-83.
- Anwar, S. (2014). *Regresi Nonparametrik Spline untuk Pemodelan Tingkat Pengangguran Terbuka*. Tugas Akhir, Institut Teknologi Sepuluh Nopember. Surabaya.
- Ariane, S. (2012). *Pendekatan Regresi Ridge Untuk Memodelkan Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka Di Provinsi Jawa Timur Dan Jawa Tengah*. Tugas Akhir, Institut Teknologi Sepuluh Nopember. Surabaya.
- Asti, F. (2011). *Regresi Multivariate pada Faktor yang Berpengaruh terhadap Pengangguran Terbuka di Jawa Timur*. Tugas Akhir, Institut Teknologi Sepuluh Nopember. Surabaya.
- Badan Pusat Statistik. (2007). *Analisis Perkembangan Statistik Ketenagakerjaan*. Badan Pusat Statistik: Jakarta.
- Badan Pusat Statistik Provinsi Jawa Barat. (2009). *Berita Resmi Statistik Keadaan Ketatanegaraan Jawa Barat*. BPS: Jawa Barat
- Badan Pusat Statistik Provinsi Jawa Barat. (2012). *Jawa Barat Dalam Angka*. Publikasi BPS: Jawa Barat.
- Badan Pusat Statistik Provinsi Jawa Barat. (2012). *Keadaan Angkatan Kerja di Provinsi Jawa Barat 2012*. Publikasi BPS: Jawa Barat
- Bellante J. (1990). *Ekonomi Ketenagakerjaan*. Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia: Jakarta.
- Bilodeau, M.L. (1992). Fourier Smoother and Additive Models. *The Canadian Journal of Statistics*, 3: 257-269.



- Budiantara, I. N. (2000). Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline. *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, 6: 285-290.
- Budiantara, I. N. dan Mulianah. (2007). Pemilihan Bandwidth Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, Vol. 10, No. 2, Hal. 159-166.
- Budiantara, I. N. (2009). Spline Dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang, *Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu: Matematika Statistika dan Probabilitas*, Institut Teknologi Sepuluh Nopember: Surabaya.
- Budiantara, I. N., Lestari, B., dan Islamiyati, A. (2010). Estimator Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Heterokedastisitas untuk Data Longitudinal, *Hibah Penelitian Kompetensi, DP2M-Dikti*: Jakarta.
- Budiantara, I. N., Ratnasari, V., Ratna, M., Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and Its Properties, *Applied Mathematical Sciences* 9(122): 6083-6094.
- Carmona, R. A. (2003). *Statistical Analysis of Financial Data in S-Plus*. Springer: New York.
- Chaubey, Yogendra P., Laib, Naamane., Li, Jun. (2011). Generalized kernel regression estimator for dependent size-biased data. *Journal of Statistical Planning Inference*, 142:708-727.
- Cox, D. dan O'Sullivan, F. (1996). Penalized Type Estimator for Generalized Nonparametric Regression, 1983. *Journal of Multivariate Analysis*, 4(56): 185-206.
- Craven, P. dan Wahba, G. (1979). Smoothing Noisy Data with Spline Functions. *Numerische Mathematics*, 31: 377-403.
- Crainiceanu, C. M., Rupper L D., dan Wand, M.R. (2004), Bayesian Analysis for Penalized Spline Regression Using Wenbugs. *Statistical Software*, 14(14): 6-14.
- De Jong, P. (1977). The Fast Fourier Transform Spectral Estimator. *Journal of Statistical Soc*, 3: 327-330.
- Eubank, R. L. (1988). *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.
- Eubank, R. L. (1999). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker: New York.
- Fan, Y. dan Liu, R. (2013). A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Models. *Tsinghua International Conference in Econometrics*

- Hardle, W. (1994). *Applied Nonparametric Regression*. Cambridge University Press: New York.
- Galtchouk, L. dan Pergamenshchikov, S. (2009). Adaptive Asymptotically Efficient Estimation in Heteroscedastic Nonparametric Regression, *Journal of the Korean Statistical Society*, 38: 305-322
- Green, P. J. dan Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall: London.
- Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press: New York.
- Hardle, W. (1994). *Applied Nonparametric Regression*. Cambridge University Press: New York.
- Hastie, T. J. dan Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall: New York, London.
- He, H. dan Huang, L. S. (2009). Double-smoothing for Bias Reduction in Local Linear Regression. *Journal of Statistical Planning and Inference*, 139: 1056-1072
- Howell, J. R. (2007). Analysis Using Smoothing Splines As Implemented in LMEO in R. Bringham Young University.
- Huang, Z. (2003). Local Asymptotic for Polynomial Spline Regression. *The Annals of Statistics*, 31 (5): 1600-1635.
- Idris dan Yanuar, R. (2007). *Ekonomi Sumber Daya Manusia*. FE UNP.
- Kayri, M. dan Zirhhoglu, G. (2009). Kernel Smoothing Function and Choosing Bandwidth for Nonparametric Regression Methods. *Ozean Journal of Applied Sciences*, 2:49-54.
- Kirn, Y.J dan Gu, C. (2004). Smoothing Spline Gaussian Regression: More Sealable Computation Via Efficient Approximation. *Royal Statistical Society, Series B*, 66 (2): 337-356
- Kusnaini. (1998). Permintaan dan Penawaran Tenaga Kerja serta Upah: Teori Serta Beberapa *Potretnya di Indonesia*. Universitas Negeri Yogyakarta: Yogyakarta:
- Liang, H. (2006). Estimation in Partially Linear Models and Numerical Comparisons. *Computational Statistics and Data Analysis*. Volume 50: L1al. 675-687
- Li, L., (2002). *Nonlinear Waveled-Based Nonparametric Curve Estimation with Consored Data and Inference on Long Memory Processes*. Midingan State University.
- Lin, Y. dan Zhang, H. H. (2006). Component Selection and Smoothing in Multivariate Nonparametric Regression. *The Annals of Statistic*, 34: 2272-2297.
- Ma, S., Yang, L. (2011). Spline-backfitted kernel Smothing of partially linier additive model. *Journal of Statistical Planning and Inference*, 141: 204-219

- Mahler. (1995). Variational Solution of Penalized Likelihood Problem and Smooth Curve. *Annal of Statistics*, 23: 1496-1517.
- Mankiw, N. dan Gregory. (2006). *Pengantar Ekonomi Makro Edisi 3*. Salemba Empat: Jakarta.
- Montoril, M. H., Morettin, P. A., Chiann, C. (2014). Spline estimation of functional coefficient regression models for tune series with Correlated errors. *Statistics and Probability Letters*, 92: 226-231
- Manzan, S. dan Zerom, D. (2005). Kernel Estimation of a Partially Linear Additive Model. *Statistics & Probability Letters*, 72: 313-322.
- Martins-Filho, C. dan Yao, F. (2009). Nonparametric Regression Estimation with General Parametric Error Covariance. *Journal of Multivariate Analysis*, 100: 309-333.
- Okumura, H. dan Naito, K. (2006). Non-parametric Kernel Regression for Multinomial Data. *Journal of Multivariate Analysis*, 97: 2009-2022.
- Prihatiningsih, O. (2012). *Menentukan Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka (TPT) Provinsi Jawa Barat dengan Regresi Terboboti Geografis (RTG)*. Skripsi, InstitutPertanianBogor. Bogor
- Qingguo, T. (2010). L1-Estimation in a Semiparametric Model with Longitudinal Data. *Journal of Statistical Planning and Inference*, 140: 393-405.
- Qu, L. (2002). *On Semiparametric Regression via Wavelets*. Dissertation. Pnrdue University.
- Rakotomamonjy, K., Mary, X. dan Canu, S. (2005). Non-parametric Regression with wavelet kernels. *Appl. Statistic Models in Business and Industry*, 21: 153-163.
- Ryan, T. P. (1997). *Modern Regression Methods*. Joh Weley and Sons: New York.
- Samsodin, M. (2012). *Regresi Spline Polynomial Truncated Multirespon untuk Pemodelan Indikator Kemiskinan di Provinsi Jawa Timur*, Tesis, Institut Teknologi Sepuluh November, Surabaya
- Santoso, N. (2009). *Klasifikasi Kabupaten Kota Di Jawa Timur Berdasarkan Tingkat Pengangguran Terbuka Dengan Pendekatan MARS (Multivariate Adaptive Regression Spline)*. Tugas Akhir, Institut Teknologi Sepuluh Nopember. Surabaya.
- Simonoff, H. S. (1996). *Smoothing Methods in Statistics*. Springer: New York.
- Sonhaji, M. (2006). Estimasi Fungsi Densitas Multivariat secara Semiparametrik. Metodologi Statistik, Badan Pusat Statistik. [http://www.batan.go.id/ppln/lokakarya/LKSTN\\_10/Mochson%20Haji-pdf](http://www.batan.go.id/ppln/lokakarya/LKSTN_10/Mochson%20Haji-pdf)
- Su, L. dan Ullah, A. (2008). Local Polynomial Estimation of Nonparametric simultaneous Equations Models. *Journal of Econometrics*, 144: 193-218.

- Sudiarsa, I. W., Budiantara, I. N., Zain, I., Suhartono, and Purnami, S. W. (2015). Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric regression. *Applied Mathematical Sciences* 9(100): 4997-5010.
- Taylor, L., W. (2009). Using the Haar Wavelet Transform in the Semiparametric Specification of Time Series. *Economic Modelling*, 26: 392-403.
- Tripena, A. dan Budiantara, I. N. (2007). *Fourier Estimator in Nonparametric Regression, International Conference on Natural Sciences and Applied Natural Sciences*. Ahmad Dahlan University: Yogyakarta.
- Wahba G. (1990). *Spline Models for Observational Data*. SIAM: Pennsylvania.
- Wang, Y. (2011). *Smoothing Splines Methods and Applications*. CRS Press Taylor & France Group: California, USA.
- Welsh, A. R dan Yee, T. W. (2006). Local Regression for Vector Responses. *Journal of Statistical Planning and Inference*, 136: 3007-3031.
- Yao, F. (2007) Asymptotic Distributions of Nonparametric Regression Estimation for Longitudinal or Functional Data. *Journal of Multivariate Analysis*, 98:40-56.
- You, J. dan Chen, G. (2007). Semiparametric Generalized Least Squares Estimation in Partially Linear Regression Models with Correlated Errors. *Journal of Statistical Planning and Inference*, 137: 117-132.

*(Halaman ini sengaja dikosongkan)*

## BIOGRAFI PENULIS



Penulis bernama singkat Rismal, dilahirkan di Belawa, Kabupaten Wajo, Provinsi Sulawesi Selatan pada tanggal 28 April 1990. Penulis merupakan anak kelima dari tujuh bersaudara dari pasangan suami-istri Bapak Husaini dan Ibu Hasnaini. Penulis menyelesaikan pendidikan formal sekolah dasar di SDN 71 Wele pada tahun 2002, kemudian menyelesaikan pendidikan sekolah menengah tingkat pertama di SMPN 2 Belawa.

Selanjutnya menyelesaikan pendidikan sekolah menengah tingkat atas di SMAN 1 Bone-bone pada tahun 2008. Pada tahun 2008 tersebut, penulis melanjutkan pendidikan ke jenjang perguruan tinggi S1 Pendidikan Matematika di Universitas Negeri Makassar (UNM) Sulawesi Selatan. Pada Agustus 2014, penulis mendapatkan kesempatan untuk melanjutkan pendidikan pada Program Magister Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) di Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Penulis dapat dihubungi di alamat email [rismal.fs@gmail.com](mailto:rismal.fs@gmail.com).

Penulis

Surabaya, Agustus 2016